Structural and biochemical characterization of carrier proteins interacting with non-elongating ketosynthase from enacyloxin polyketide synthase

by

Alma Svatoš

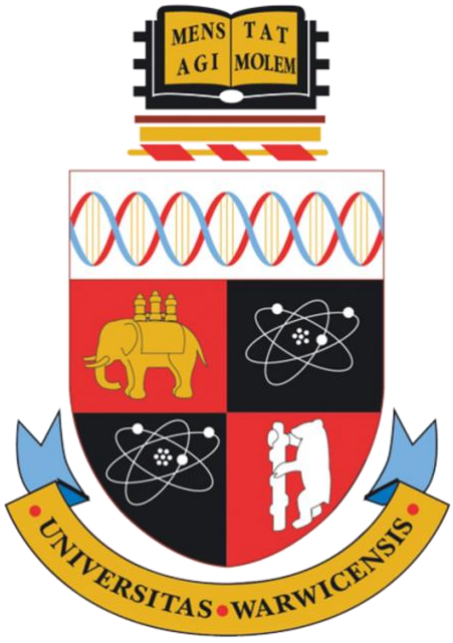
Thesis

Submitted for the degree of

Doctor of Philosophy

The University of Warwick

September 30, 2019



Contents

Acknowledgements

I would like to thank my supervisors; Prof. Józef Lewandowski and Prof. Greg Challis for their continuous guidance and support during my PhD. I am particularly grateful to them for taking the time to meet with me on a regular basis and discuss my work.

I would also like to thank Dr. Angelo Gallo for sharing a wealth of his NMR knowledge with me and helping me setting up solution and solid-state NMR experiments. He also provided inputs to my PhD thesis. Dr. Simone Kosol dedicated time to help me setting up in the lab when I was starting my PhD and was always willing to provide her extensive knowledge of biochemistry to guide my work. She also created short Bamb_5919 ACP construct that was the basis of my project. Dr. Joleen Masschelein helped me setting up bioactivity assays and also ran the LC-MS analysis of my proteins samples. I am also grateful to Dr. Trent Franks for the help with solid-state NMR experimental setup.

I would further like to thank Dr. Antonio Laezza and Dr. Sarah-Jane Smith for the assistance with setting up BLI experiments and prof. M. Gibson who gave me the access to use the Octet™ BLI machine. I appreciate the help from Dr. Nikola Chmel with the analysis of CD experiments and Dr. Arun Gupta who ran ITC experiments of my samples.

Dr. Lona Alkhalaf provided useful suggestions regarding my initial thesis draft.

Prof. Ann Dixon and Dr. Cristophe Corre, my advisory committee, provided insightful comments and encouragement.

It was a pleasure being a part of both Lewandowski and Challis groups and enjoying friendly and supportive atmosphere both inside and outside the lab and office.

Finally, my PhD years would not have been the same without a wealth of support from my parents Vesna and Vladimir, sister Alenka and fiancée Matija.

My PhD project was funded by the ERC Starting Grant 639907 provided to Prof. J. R. Lewandowski.

Declaration

The thesis is a summary of research performed at the Department of Chemistry at the University of Warwick, between October 2015 and September 2019.
It is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented, including data analysis, is wholly on my own, except in the cases outlined below:

- LC-MS samples were run by Dr. Joleen Masschelein
- ITC samples were run by Dr. Arun Gupta.

Abbreviations

A	Adenylation Domain
ACP	Acyl Carrier Protein
AMR	Antimicrobial Resistance
ARO	Aromatase
AT	Acyl Transferase
BLI	Bio-layer Interferometry
C	Condensation Domain
CD	Circular Dichroism
CLF	Chain Length Factor
CoA	Coenzyme A
CP	Cross Polarization
CSA	Chemical Shift Anisotropy
Cy	Cyclisation Domain
CYC	Heterocyclization Domain
DEBS	6-Deoxyerythronolide B Synthase
DH	Dehydratase
DHCCA	(1S, 3R,4S)-3,4-dihydroxycyclohexane Carboxylic Acid
DNA	Deoxyribonucleic Acid
DSS	Sodium trimethylsilylpropanesulfonate
EF-Tu	Elongation Factor Thermo Unstable
ER	Enoyl-acyl Carrier Protein Reductase
FAS	Fatty Acid Synthase
FPLC	Fast Protein Liquid Chromatography
Gd(DTPA-BMA)	Gadolinium Diethylenetriaminepentaacetic Acid Bismethylamide
Gfp	Green Fluorescent Protein
GTP	Guanosine-5'-triphosphate
TOCSY	Total Correlated Spectroscopy
HMQC	Heteronuclear Multiple Quantum Coherence
HSQC	Heteronuclear Single Quantum Coherence
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer

ITC	Isothermal titration calorimetry
KR	Ketoreductase
KS	Ketosynthase
KS^0	Non-elongating Ketosynthase
LB	Lysogeny Broth
MAS	Magic Angle Spinning
MAT	Malonyl-CoA:ACP Transacylase
MRSA	Methicillin-resistant *Staphyloccocus aureus*
MT	Methyltransferase
MWCO	Molecular Weight Cut-off
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhouser Effect
NOESY	Nuclear Overhouser Effect Spectroscopy
NRPS	Nonribosomal Peptide-Synthetase
OD_{600}	Optical density of a sample measured at a wavelength of 600 nm
Ox	Oxidase
PCP	Peptidyl Carrier Protein
PDB	Protein Data Bank
PKS	Polyketide Synthase
PPTase	Phosphopantetheinyl Transferase
PTM	Post-translational Modification
RPM	Revolutions Per Minute
SDS-PAGE	Sodium Dodecyl Sulfate-polyacrylamide Gel Electrophoresis
SEC	Size-exclusion Chromatography
Sfp	Surfactin Phosphopantetheinyl Transferase
SLiM	Short Linear Motif
SPR	Surface Plasmon Resonance
TE	Thioestherase
TEV	Tobacco Etch Virus
TOCSY	Total Correlated Spectroscopy

UHPLC-ESI-Q-TOF-MS	Ultra-high-performance Liquid Chromatography Coupled With Electrospray Ionization Quadrupole Time-of-flight Mass Spectrometry
[U-^{13}C]	Uniformly Labelled With ^{13}C Isotopes
[U-^{13}C, ^{15}N]	Uniformly Labelled With ^{13}C And ^{15}N Isotopes
WHO	World Health Organisation

Abstract

Enacyloxin is an antibiotic synthesized inside bacterial cells by gigantic modular multienzyme complex called polyketide synthase (PKS), which is active against multidrug-resistant *Acinetobacter baumannii* that leads the World Health Organization 'priority pathogens' list. In spite of its promising physiological activity, enacyloxin IIa has not been used as a drug due its structural instability. The modular nature of enacyloxin and other PKSs makes them amenable to rational engineering to elucidating protein-protein interactions within PKSs will be crucial for their bioengineering to produce novel drugs.

A key step in enacyloxin biosynthesis is the transfer of the polyketide chain from an acyl carrier protein (ACP) domain in the last PKS module to a standalone peptidyl carrier protein (PCP) by a non-elongating ketosynthase (KS^0) domain. It has been reported that this transfer is necessary for the condensation (C) domain to catalyse polyketide chain release via intermolecular esterification with 3, 4-dihydroxycyclo-hexane carboxylic acid (DHCCA). No esterification occurs when the substrate is attached to the ACP domain. Thus, the KS^0 transfers the substrate from a carrier protein that is unable to interact with the C domain to one that it is able to interact with.

An experimental approach combining NMR and biochemical techniques presented in the thesis suggests that the length of helix 1 and the lack of charged residues in the C-terminal part of ACP prevent it from establishing favourable interaction with C domain.

Chapter 1. Introduction

1.1 Natural products

Natural products are substances of a great structural variety and with a molecular weight normally below 3000 Daltons (Figure 1.1)[1].

Within the field of organic chemistry, natural products are usually defined as purified organic compounds isolated from natural sources that are produced by the primary or secondary metabolic pathways[2]. However, within the field of medicinal chemistry, the definition is often restricted to secondary metabolites[3]. Unlike primary metabolites, secondary metabolites are not necessary for an organism to live, grow, develop or reproduce. Rather, they are produced as a consequence of the organism adapting to the surrounding environment to act as a mechanism of defence against predators and to compete for space and nutrients[4].

Natural products and their derivatives have been used as a source of therapeutic remedies for millennia, dating back to ancient Mesopotamia. However, the first antibacterial microbial natural product, penicillin, was discovered by Alexander Fleming in 1928. The period that followed was a golden age of drug discovery which resulted in the discovery of many natural products and their development into drugs[5] (Figure 1.1). The introduction of antibiotics such as tetracycline and kanamycin has significantly decreased the mortality rate from bacterial diseases.

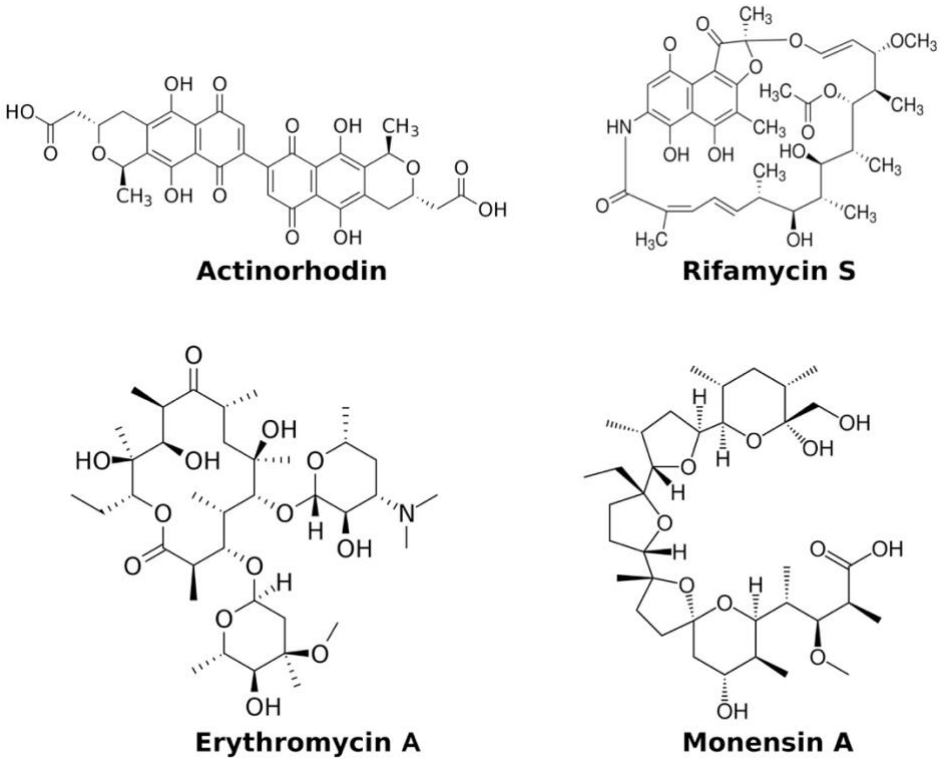


Figure 1.1 Examples of natural products with antibacterial properties.

Although natural products have been crucial in drug discovery, pharmaceutical companies have either terminated or substantially reduced their natural product research in the past few years[6]. The reasons for this were the progress in metagenomics and combinatorial chemistry. Moreover, not all natural products can be completely synthesized due to their structural complexity. The synthesis of such compounds on an industrial scale would be expensive and time consuming.

However, with the help of advances in drug discovery and increased medicinal chemistry knowledge, pharmaceutical and biotechnology industries are once again turning to natural products to harness new useful derivatives of these therapeutics[7]. This is often driven by the therapeutic need to treat complicated conditions such as cancer and infections initiated by multi-resistant pathogens. Moreover, the treatments against an increasing number of multi-resistant pathogens require more effective drugs.

1.2 Antimicrobial resistance (AMR)

The misuse and overuse of antibiotics in farming and agriculture has allowed pathogenic bacterial strains to acquire resistance, which has become an increasing problem in recent years. At the same time, the pace at which novel antibiotics are being discovered has reduced considerably. It is not just a problem limited to bacteria, but all microbes that have the potential to mutate and make drugs ineffective. The progress made over the past few decades to manage diseases such as malaria and HIV could be reversed, with these conditions once again spiralling out of control. It is thought that over 70% of pathogenic bacteria are resistant to at least one antibiotic. For instance, in 15 European countries, more than 10% of bloodstream *Staphylococcus aureus* infections are caused by methicillin-resistant strains (MRSA), with several of these countries seeing resistance rates closer to 50%[8]. Also, colistin, discovered in the 1940s, was rarely used due to many negative side-effects and remained an effective last resort antibiotic. However, over the past few years, numerous cases of genes that confer colistin resistance through horizontal gene transfer have been found, including cases in the UK and US[9]. Transduction, a horizontal gene transfer mechanism mediated by bacteriophages that infect bacteria, is thought to be the primary means by which bacteria acquire resistance to antibiotics.

With globalisation increasing the spread of the disease, it is now imperative to produce novel drugs. This has led to the World Health Organisation (WHO) categorising AMRs as a severe threat to global public health.

1.3 Enacyloxin IIa

One of the antibiotics that could overcome current antimicrobial resistance is enacyloxin IIa, which is synthesized by *Burkholderia ambifaria,* a bacterial strain which can cause severe lung infections in people with cystic fibrosis[10]. Enacyloxin IIa is effective against *Acinetobacter baumannii*, known as one of the most difficult multidrug-resistant gram-negative bacilli to control[11]. *Acinetobacter* species is associated with nosocomial infections including pneumonia, septicemia, wound sepsis, urinary tract infection,

endocarditis, and meningitis[12]. *A. Baumannii* has become an emerging pathogen especially in the hospitals due to its ability to survive in adverse environmental conditions[13].

Therefore, it is categorised as a priority pathogen on the 2017 WHO's list of bacteria for which new antibiotics are urgently needed. Enacyloxin IIa is a polyene polyketide, which is a class of natural products that are synthesized by a polyketide synthase biosynthetic gene cluster (Figure 1.2). Polyketide synthases are described later in the chapter. Unlike usual polyenic antibiotics that break down the cell membranes of yeast and fungi, enacyloxin IIa interferes with protein synthesis by targeting elongation factor Tu (EF-Tu), a crucial mediator of translation at the ribosome[14]. The prokaryotic EF-Tu binds aminoacylated tRNA and moves it onto a free site on the ribosome to associate with the mRNA, followed by guanosine triphosphate (GTP) hydrolysis into guanosine diphosphate (GDP) if the match is correct. Bound to ET-Tu GTP, enacyloxin IIa inhibits EF-Tu release, thus blocking the transfer of polypeptide-tRNA complex of the ribosomal P site onto the adjacent A-site[15] (Figure 1.2).

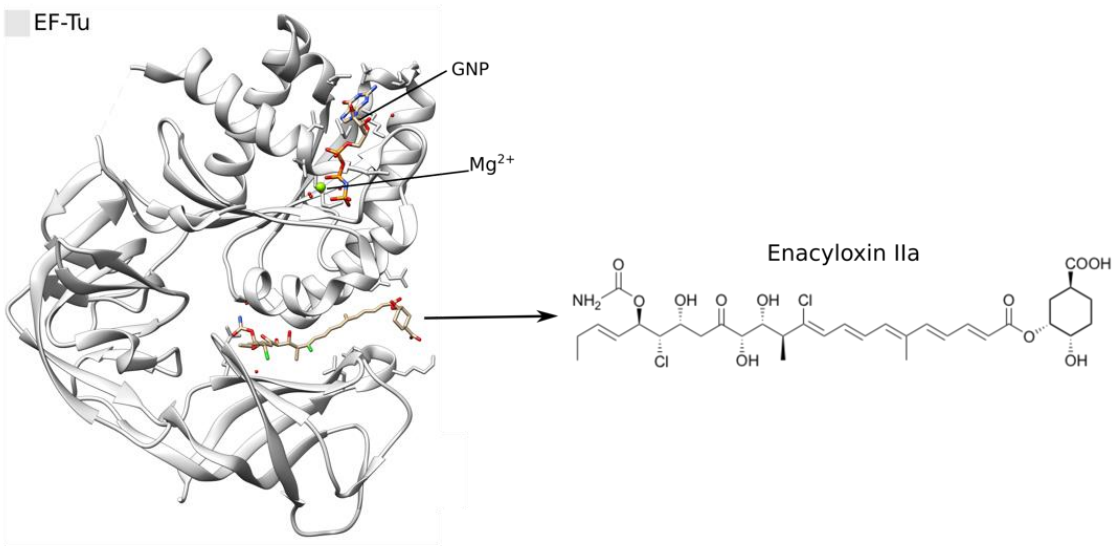


Figure 1.2 Enacyloxin IIa – the structure and mode of action. The figure shows the crystal structure of the Escherichia coli EF-Tu·guanylyl iminodiphosphate (GDPNP)·enacyloxin IIa complex and the bound enacyloxin IIa. Elongation factor (EF-) Tu·GTP transfers aminoacyl-tRNA to the ribosome. Enacyloxin IIa hinders bacterial protein synthesis by blocking the release of EF-Tu·GDP (GNP) from the ribosome[15] (PDB:2BVN).

In spite of its favourable antimicrobial properties, enacyloxin IIa is not stable enough to make a good drug candidate. For example, the ester bond (Figure 1.2) can be easily hydrolysed in the body and unstable conjugated double bonds are complicated to isolate. Besides, its complex structure renders the *in vitro* synthesis difficult. However, using synthetic biology approaches, it might be possible to manipulate enacyloxin IIa biosynthesis to create pharmaceutically useful products.

1.4 Polyketide synthases and non-ribosomal polyketide synthases

Enacyloxin IIa is biosynthesized by a giant bacterial multi-enzyme complex which is a hybrid of a large biomolecular complex called polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS). The enacyloxin synthetic pathway is described later in this chapter. Polyketides are a family of bioactive natural products that have found widespread commercial application in pharmaceutical and agrochemical industries. Polyketides are biosynthesised by enzymes called polyketide synthases (PKSs) that are organised as a series of modules comprised of linked domains. Due to their modular organisation, PKSs represent targets for bioengineering, enabling the synthesis of functionally optimised natural products.

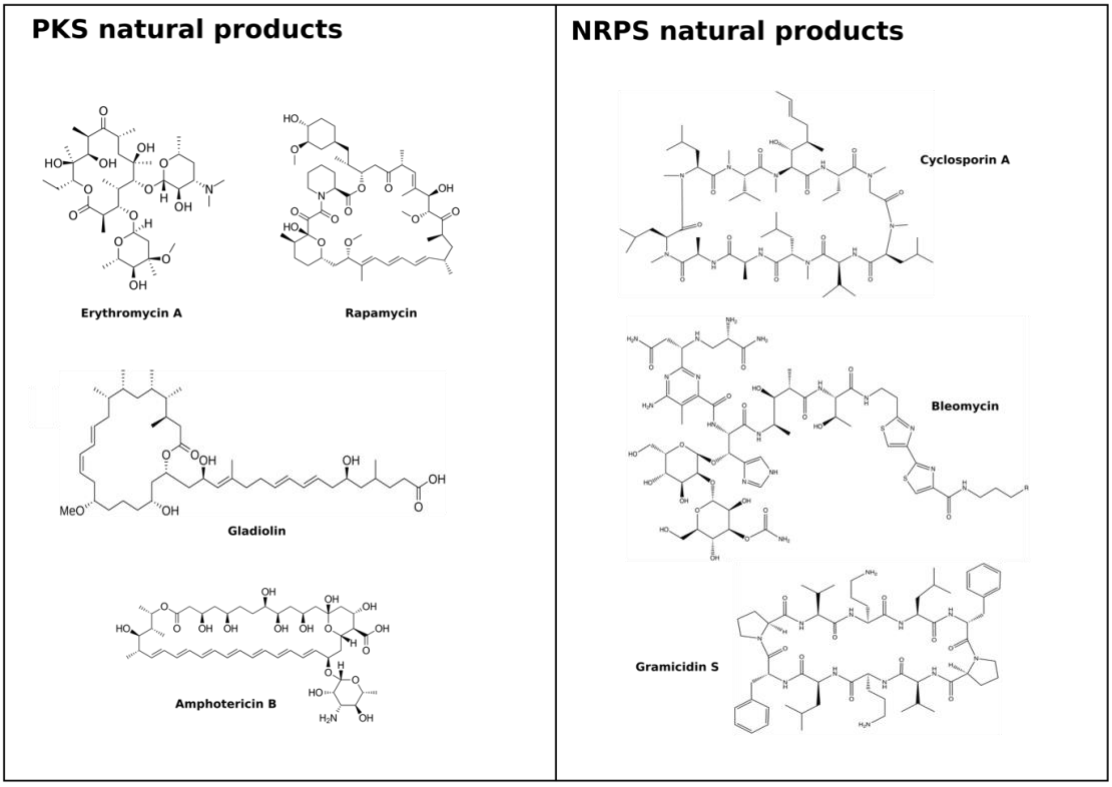


Figure 1.3 Examples of natural products produced by bacterial PKSs.The natural products produced by NRPSs differ from PKS synthesised products as they are synthesized from proteinogenic and non-proteinogenic amino acids. Some of the compounds presented in the figure are used clinically as antibiotics (erythromycin, amphotericin B,gramicidin), lipid-lowering drugs (lovastatin), immunosuppressants (cyclosporin) and to treat cancer (bleomycin). Gladiolin has a promising activity against Mycobacterium tuberculosis.

1.4.1 Polyketide biosynthesis

The main building units of polyketides are acyl-CoAs. Their biosynthetic scheme resembles that of fatty acid synthases (FASs)[16]. In both polyketide and fatty acid biosynthesis, the growing chain is covalently attached to an acyl carrier protein (ACP) via a phosphopantetheine (PPant) arm[17,18]. The PPant, derived from CoA, is attached post-translationally to a conserved serine residue on the *apo*-ACP by a phosphopantetheine transferase (PPTase) to yield *holo*-ACP (Figure 1.4.)[19,20]. Conversion from apo to holo-ACP is important for the bioactivity of ACP as the newly synthesized ACP is in the form of apo-ACP, which is a non-functional form of the protein that cannot transfer the substrate and has to be converted to the functional holo-ACP[21,22].

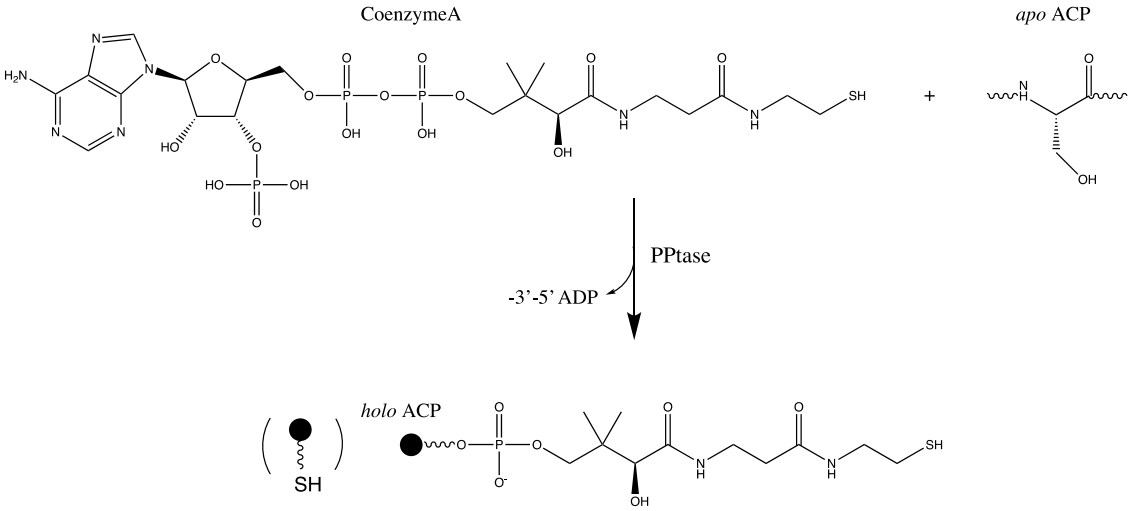


Figure 1.4 Conversion from apo-ACP to holo-ACP. Catalysed by a PPTase, phosphopantetheine is posttranslationally attached from co-enzyme A to a conserved serine residue on an apo-ACP.

The *holo*-ACP, together with an acyltransferase (AT) and a ketosynthase (KS), comprise the minimal set of enzymatic domains required for a single chain elongation reaction, also being called a module (Figure 1.5). Polyketide biosynthesis is initiated by the loading of a starter unit, an acyl-CoA derivative, onto the Ppant arm of the ACP.

In fatty acid biosynthesis, the starter unit is typically an acetyl, whereas PKSs can use several variants such as acetyl-, propionyl- and butyryl[23]. AT domain catalyses the transfer of a malonyl-derived extender unit from a CoA-thioester onto the thiol of the ACP phosphopantetheine moiety[24]. The elongating carbon-carbon bond is then formed by Claisen condensation, which is catalysed by the KS domain (Figure 1.5.). Further modifications are generated at the β-keto position by ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains, to form β-hydroxyl and fully saturated intermediates [23].

The growing polyketide chain is transferred from the ACP to the KS of the downstream module, where further elongation and modification is carried out. The polyketide chain is therefore shifted from module to module until the final polyketide structure is assembled. Following the last elongation cycle, typically, a thioesterase (TE) domain catalyses hydrolysis or lactonisation, which releases the polyketide chain from the PKS[25].

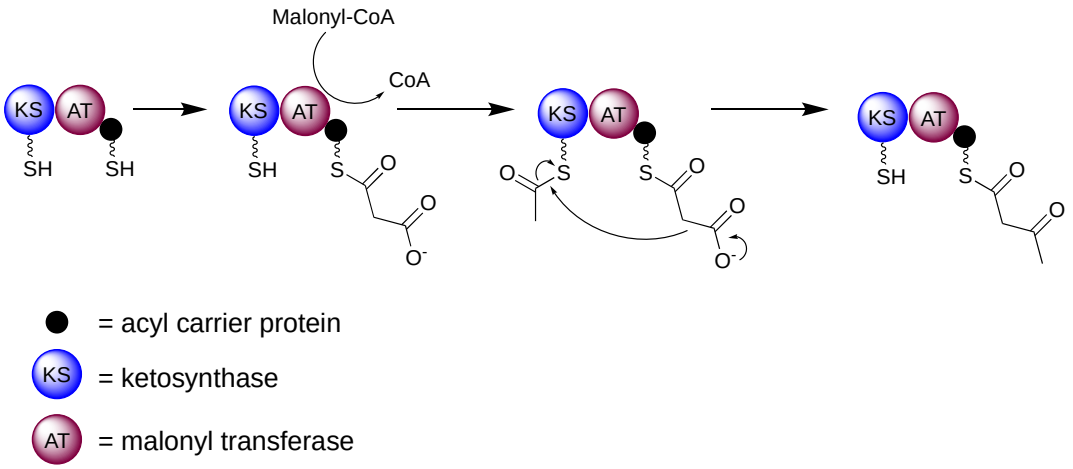


Figure 1.5 The mechanism of polyketide chain formation. The enzymatic domains required for a single cycle of chain elongation are KS, AT and ACP (acyl carrier protein) ('minimal module'). AT domains loads malonyl-derived precursors onto the thiol of the ACP.

The released polyketide chain is usually only a precursor to the bioactive compound. Tailoring enzymes are required to produce the final product by adding modifications, including glycosylation, methylation and hydroxylation[23].

1.4.1.1 Polyketide synthase classification

PKSs can be divided into three classes based on their modular architecture and mode of action[26,27]. Type I PKSs are primarily found in bacteria and fungi, whereas type II PKSs only in bacteria. Type III PKSs are mainly observed in plants. The type I PKSs can be sub-divided into the iterative type I PKSs, which are found in both bacteria and fungi, and the modular type I PKSs can be only found in bacteria.

Iterative Type I PKSs

Iterative type I PKSs make use of a single module to carry out elongation steps. β-keto processing is dictated by the reductive domains within the module, generating non-reducing (NR), partially reducing (PR) and highly reducing (HR) products. Despite being restricted to a single module, the level of reduction may not be the same for each elongation step[28].

Modular Type I PKSs

Modular type I, or non-iterative PKSs, are comprised of multiple sequential modules which produce the polyketide product. Each module catalyses a single cycle of chain elongation[28]. A complete modular type I PKS complex usually consists of two or more large polypeptide chains which leads to overall molecular weights on the MDa scale. The 6-deoxyerythronolide B synthase (DEBS) system, responsible for the biosynthesis of an antibiotic erythromycin A (the chemical structure is shown in Figure 1.2), is a prototypical example of such a PKS[29,]. Here, the biosynthesis is initiated by the loading of a propionyl unit onto the first ACP by a loading module consisted of an AT and an ACP. The subsequent AT domains load methylmalonyl extender units onto the ACPs for chain extensions. After six successive chain elongations, a TE domain catalyses the macrolactonisation of the polyketide chain, which results in the release of the product 6-deoxyethronolide B, which is a precursor of erythromycin (Figure 1.6).

The modular nature of PKSs and sequential modifications of the intermediates in modular type I PKS systems allow prediction of the final product from the understanding of the domain order. This relation between domain architecture and the final product is referred to as colinearity[17]. By understanding domain organisation, it was possible to construct biosynthetic schemes of several PKSs just from genetic sequences[31,32].

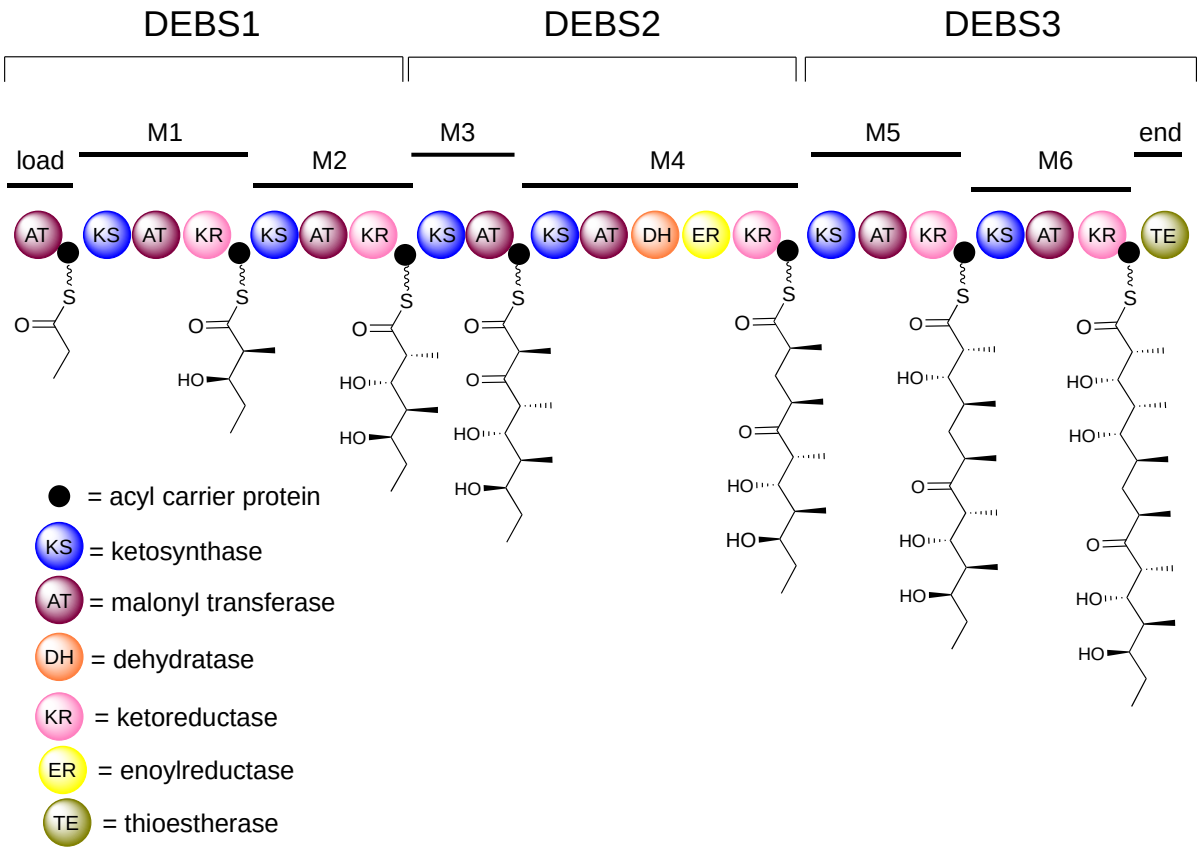


Figure 1.6 Modular organisation of a prototypical PKS system – DEBS. The biosynthetic gene cluster of Saccharopolyspora erythraea contains three large open reading frames – DEBS1, DEBS 2, DEBS3, which code for three polypeptides[30]*. Each DEBS polypeptide is comprised of functional domains that are shown are shown, along with intermediate polyketide chains produced. Stepwise synthesis of 6-dEB begins at DEBS1 and ends with cyclization by the TE domain to yield 6-dEB, which is further functionalized (not shown) to yield erythromycin.*

Cis- and *trans*-AT PKSs

Cis- and *trans*- class of modular type I PKSs do not follow collinearity rule since their domain organisation does not always correlate with the product due to the high diversity of modules that often contain novel enzymatic domains[33]. The most characteristic feature of *trans*-AT PKSs is the absence of AT domains whose activity is provided by free-standing domains servicing every module in the entire cluster[34]. In *cis*-AT PKS, the AT domains are integral to the PKS, with an AT domain within each module.

Type II PKSs

Type II PKSs adopt an iterative mechanism. However, unlike iterative type I PKSs, their catalytic domains are expressed on different genes. Type II PKSs are composed of separate functional enzymes that interact during synthases to form a polyketide structure. The polyketide structure is then enzymatically converted to the cyclic form to produce polycyclic aromatic compounds such as an antibiotic tetracycline and a chemotherapy drug doxorubicin.

Characteristic of type II PKS group is the core module of subunits called 'minimal PKS', which is comprised of an ACP, a malonyl-CoA:ACP transacylase (MAT) and two KS domains; KS_α and KS_β, where KS_β dictates the length of the polyketide chain[35]. For that reason, KS_β is called chain length factor (CLF). The final product is formed by the additional domains including ketoreductases (KR), cyclises (CYC) and aromatases (ARO).

Type III PKSs

The type III PKSs have the simplest domain structure which includes homodimers of relatively modest-sized proteins (40–45 kDa). They catalyse the assembly of polyketides by iterative decarboxylative malonyl-CoA condensations, in a biosynthetic process that closely resembles biosynthesis of fatty acids, but without the presence of acyl carrier proteins (ACPs). The chain elongation of a CoA-tethered starter molecule is typically accompanied by cyclization of the linear intermediate to create different polyketide scaffolds[36–38].

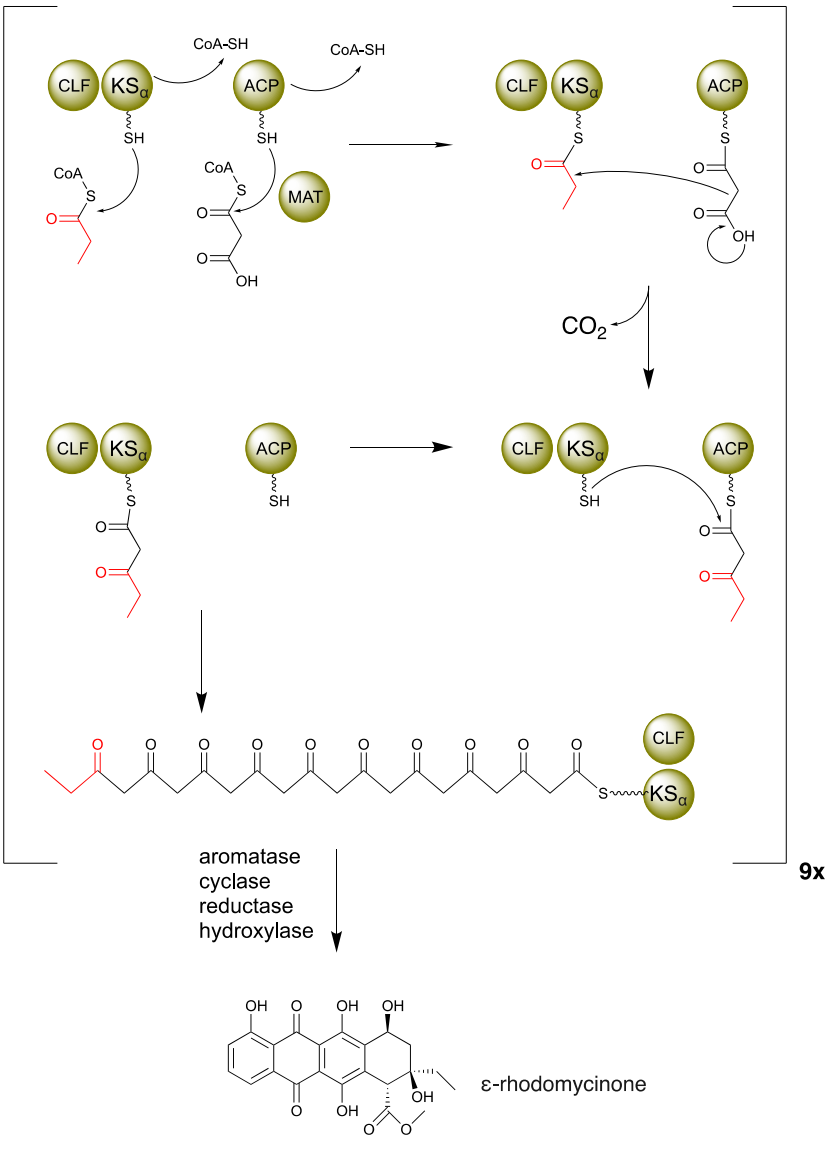


Figure 1.7 *Biosynthesis of rhodomycin, an example of an iterative type II PKS. The figure shows a mechanism of 2-carbon decaketide chain synthesis to produce* ε-rhodomycinone, *the doxorubicin precursor. The doxorubicin PKS condenses one propionyl-CoA unit and nine malonyl-CoA units. Initially, KS, CLF, and ACP catalyze a formation of a diketide between propionyl-CoA and malonyl-CoA. The same domains catalyze eight more condensation reactions, starting with the previous diketide and adding successive malonyl units to produce a decaketide*[34].

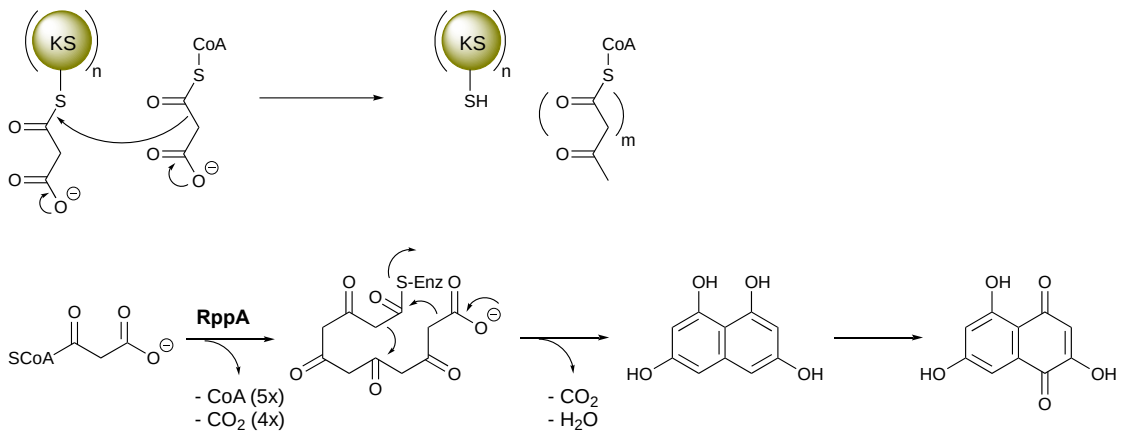


Figure 1.8 Schematic representation of a type III biosynthetic mechanism, as exemplified by the RppA synthase for the biosynthesis of aromatic polyketide flavolin.[39] *The figure was reproduced from Shen, 2003.*

1.4.1.2 Structure and mechanism of PKS domains

Core PKS domains are mechanistically and structurally relatively well understood so their structural features will be discussed below.

The ketosynthase domain

The ketosynthase (KS) domain catalyses the carbon-carbon bond forming thioester-templated Claisen condensation reaction required for the chain elongation step in polyketide and fatty acid biosynthesis[40]. Catalysis is initiated when an intermediate is transferred through transthioesterification from the ACP of one module to the KS of the next module. An ACP with an attached extender unit then binds to the polyketide-bound KS to allow for a decarboxylative condensation. This generates a β-keto intermediate that can be processed at its α- and β-carbons by methyltransferase (MT), ketoreductase (KR), dehydratase (DH), enoylreductase (ER), or β-branching enzyme[34].

Crystal structures of KS domains from both FASs and PKSs exhibit a highly conserved 2α-5β-2α-5β-2α thiolase fold (Figure 1.9.A and B)[41,42]. KS domains are generally dimeric, and their interface regions contribute significantly to the dimerisation of the type I PKSs. The majority of KS domains have a Cys-His-His catalytic triad active site[40] (Figure 1.9.C).

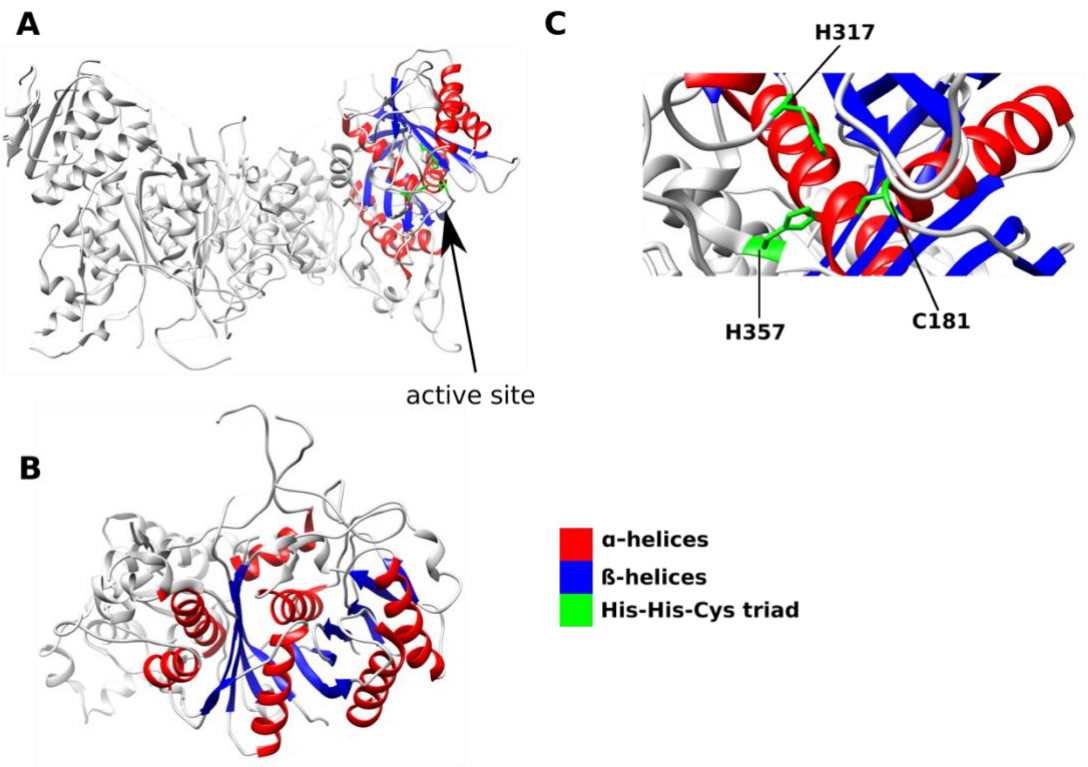


Figure 1.9 Structural analysis of the ketosynthase from module 1 of the bacillaene synthase[43] *(PDB: 5ELP). A) KS dimer with a highlighted 2α-5β-2α fold and and active site. B) Close-up of the 2α-5β-2α fold. C) A His-His-Cys catalytic triad is typical of KS from most organisms.*

The insights into both processes have been obtained from crystal structures of a FAS KS and *E.coli* FabF with platensimycin[44,30]. These structures suggested that the oxyanion hole stabilises the acyl-KS. Oxyanion hole is formed by the amide hydrogens of the active site Cys and a conserved Phe, which is believed to mediate the acylation of the active site Cys[45]. The crystal structure of platesimycin-FabF complex revealed that the carbonyl of the glutamine side chain was hydrogen-bonded to the two backbone amides in the oxyanion hole[46]. This led to conclusion that only one of the His residues is catalytically important in the activation of the carboxylate extender unit. Another histidine is thought to act as a base which activates a water molecule to facilitate nucleophilic attack on carboxylate[47]. It is believed that the coordination of the carbonyl in the oxyanion hole facilitates the organisation of active site residues to perform decarboxylation and elongation[48].

The KS mode of action (Figure 1.10) involves acylation where the thioesterification starts when an acyl-ACP enters the KS substrate-binding. There is also a phosphopantetheine-

binding area inside KS. The amino acid sequence of phosphopantetheine-binding region is well conserved, but the structure of the substrate-binding area has not been yet characterised. The pK_a value of the active site cysteine is low for a thiol due to its positive charge[34] which increases the nucleophilicity of its thiol, enabling attack of the carbonyl. The resulting negative charge on the thioester carbonyl is stabilised by an oxyanion hole. The intermediate then releases the Ppant-ACP, which leaves the growing polyketide chain attached to the KS. The next step involves the decarboxylation of the malonyl or methylmalonyl chain extender unit that is supplied by the upstream ACP. Upon binding, a water molecule that is activated by one of the catalytic histidine residues facilitates the attack at the malonyl carboxylate, which releases $HCO3^-$. The resulting enolate is stabilised by a conserved histidine residue. The last condensation step involves the enolate-mediated attack on the acyl-thioester. The resulting tetrahedral intermediate is stabilised by the oxyanion hole. This is followed by the formation of a β-keto product. The elongated intermediate is linked to the receiving ACP and is diffused out of the active site.

Non-elongating ketosynthase domains

The variants of KS are sometimes employed by the domains to fulfil specific functions within the biosynthetic cluster, such as the non-elongating KS domains, called KS^0. These domains lack a crucial conserved active site histidine, which is required for chain elongation[49]. It is believed that they function only as a shuttle to pass the intermediates onto the next module or to allow modification from a downstream domain without any chain elongation occurring[19].

However, non-elongating ketosynthases play important roles in PKSs by catalysing translocation of acyl chain intermediates[50,51].

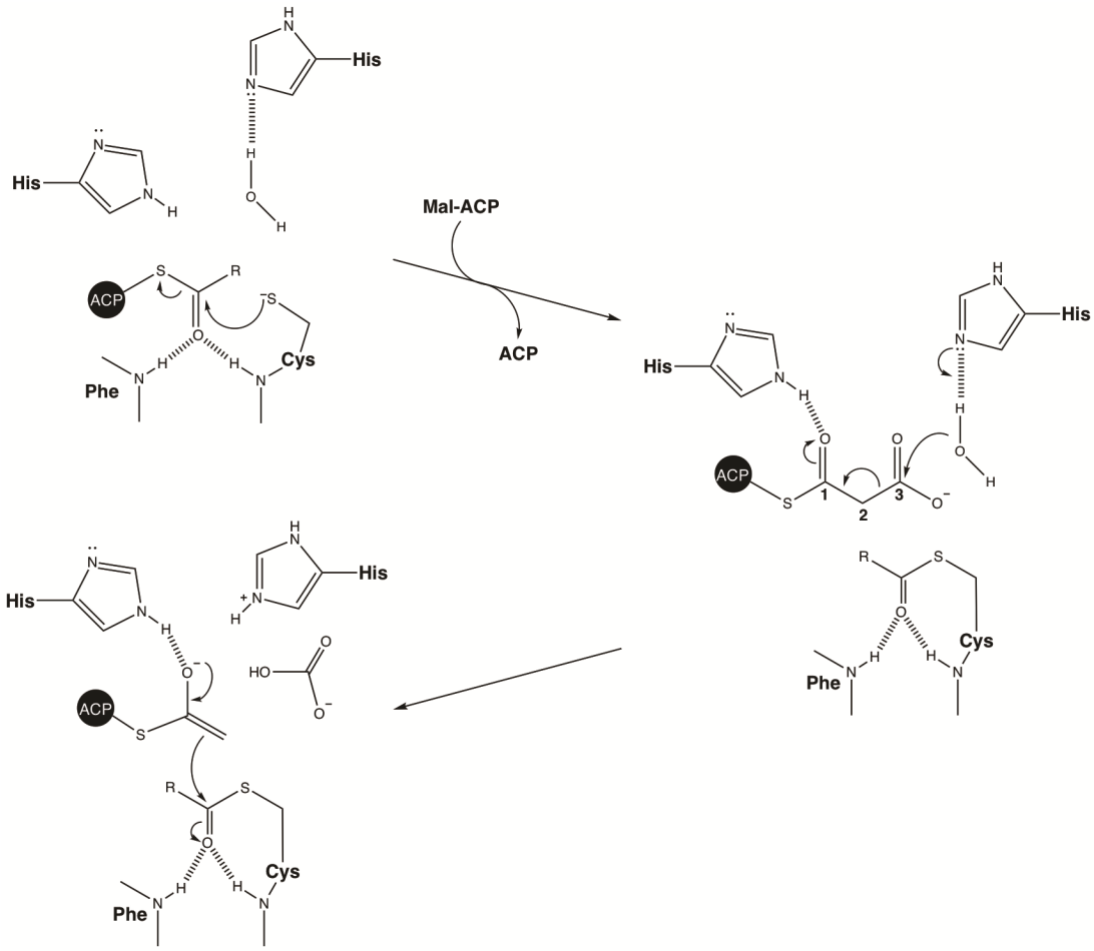


Figure 1.10 *Schematic representation of a proposed mechanism for the ketosynthase mode of action. The initial step involves the transfer of the acyl group to the acyl-ACP active site cysteine. Backbone amides of Phe and Cys form the oxyanion hole, which neutralizes the thioester carbonyl negative charge. Malonyl-ACP binds to the enzyme upon the release of ACP. Histidine activates a catalytic water molecule to release bicarbonate by attacking the malonyl-ACP carboxylate. Finally, β-ketoacyl-ACP product is formed. The figure was reproduced from Zhang, Y. et al, 2006.*

The acyl carrier protein

Acyl carrier proteins (ACPs) typically consist of four α-helices connected by three loops that are stabilised by a hydrophobic core (Figure 1.11.)[52–55]. The conserved serine, located at the N-terminal side of helix II, is post-translationally modified to incorporate an ~18Å long phosphopantetheine chain. Phosphopantethiene moiety is an essential prosthetic group of ACP, as it mediates covalent attachment of the intermediates to the

PKS via a thioester linkage thus providing the required flexibility to reach active sites of other domains within a given module[56].

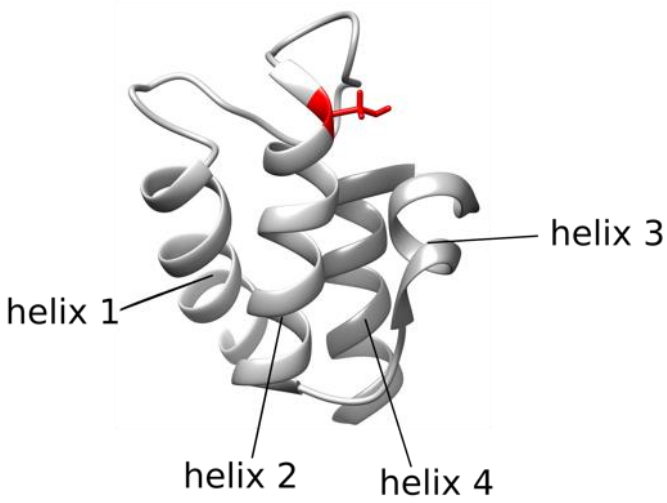


Figure 1.11 The NMR structure of B. subtilis apo-ACP (PDB: 1HY8). Conserved serine residue is highlighted in red.

Specific type I PKS ACPs (EryACP2) have a short N-terminal helix (helix 0) and it is believed to facilitate the docking to downstream KS domains[19]. ACPs transfer acyl groups to the KS domains that are immediately downstream, which implies that interactions between helix II of ACPs and KS domains are highly specific[57]. Crystal structures of ACPs from type II FAS indicate that the fatty acid acyl chain binds between the helices of the ACP domain[19]. This substrate sequestration can lead to minor alterations in ACP structure due to the perturbation of the helical core of the ACP by the substrate[58]. This is not believed to be the case for type 1 ACPs, as suggested by the NMR studies of the acyl-ACP from the curacin cluster[19].

Protein-protein interfaces within type I PKS subunits

The crystal structures of KS-AT didomain, DH, ER-KR and a TE, together with NMR structures of ACPs provided details about the boundaries between domains within PKSs[19,59–63].

In type I PKSs, the dimerization of the modules is mediated by the KS, which has a large dimeric interface in comparison to that of ACPs[64]. The other domains that commonly

establish an intermodular contact are the DH and TE. However, they also have a smaller interface than the KS domains[64]. AT, KR, ER and ACP are primarily monomeric. This allows the flexibility for the conformational changes required for PKS functions.

In PKSs the dimers are often stabilised by dimerization motifs located at the ends of the modules. The motifs that facilitate KS and DH dimerisation include a coiled-coil that is frequently located at the N-terminal end of a PKS module and a four-helix bundle often present at the C-terminus.[65–68].

Structure of the Type I PKS complex

In recent years, the understanding of type I PKSs structure has significantly improved with the progress in electron microscopy. EM revealed the quaternary structure of the dimeric metazoan[69] and porcine[70] FAS. The first details about KS-AT and KR interdomain boundaries were obtained from erythromycin PKS[62]. Although the type I PKSs share a large amount of structural similarity with the FAS, there are some differences between them. Compared to the dimeric FAS ER, the ER domain in type I PKSs is monomeric. These domain-domain structures of PKSs were generated by using NMR and X-ray crystallography so they have limitations due to differences in the oligomeric state and structure of individual catalytic domains. However, the recent advances in cryo-EM technology led to the first structural insights into the catalytic rearrangements of an intact PKS module. The models of the last module of pikromycin PKS (PikAIII) were acquired for each conformation of its catalytic cycle[71]. The study showed that KS, AT, and KR domains form a chamber. The dimeric *apo*-PikAIII was observed in two conformations. In the first conformation, ACP was located inside the chamber and interacted with the KR domain (Figure 1.12 1A, B). In the second conformation, it was near the KS and AT domains (Figure 1.12 2A, B). The removal of ACP resulted in a rotation of the KR domains around the AT domains which suggested that ACP plays a role in the positioning of the KR relative to the AT domain. Also, the study demonstrated that the upstream ACP binds to the KS domain outside PikAIII's chamber. Furthermore, loading of a methylmalonyl extender unit onto PikAIII resulted in the ACP relocating to a second interaction site on the KS, this time inside the PikIII chamber. This provided evidence that the KS has two

different entrances for substrate loading from the upstream ACP and offloading to the intramodular ACP.

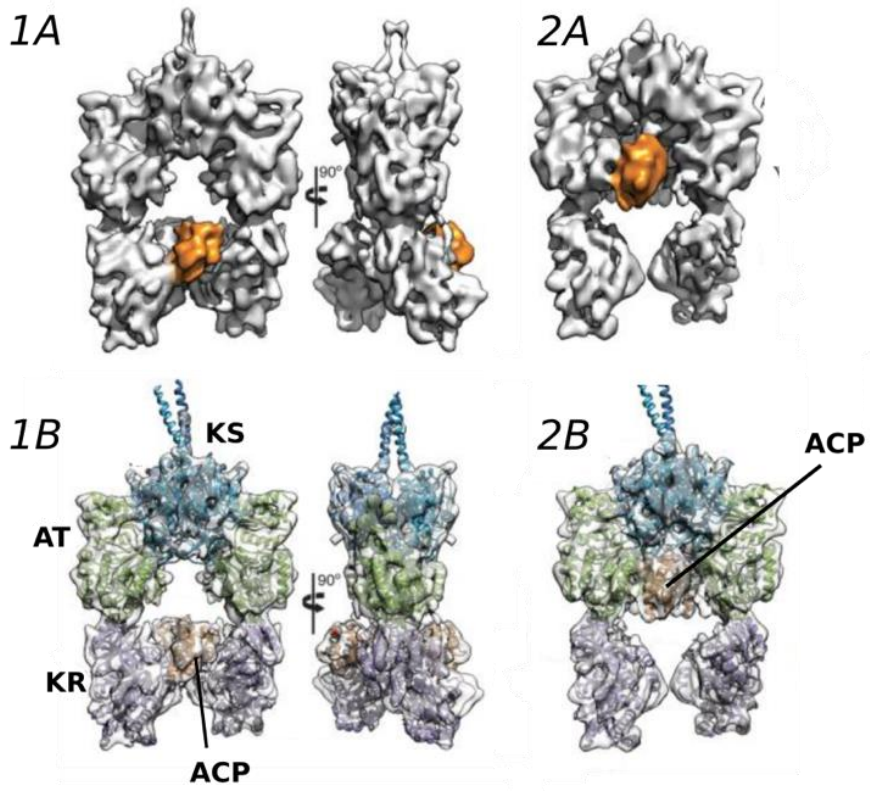


Figure 1.12 Cryo-EM structures of holo-PikAIII[71]*. Solid (1A) and transparent (1B) representations with modelled structures of the cryo-EM map of holo-PikAIII conformation, in which the ACP_5 (orange) is near KR_5. Solid (2A) and transparent (2B) representations with moddele structures of the cryo-EM map of holo-PikAIII conformation in which ACP_5 (orange) is near AT_5. The figure was reproduced from Dutta et al., 2015.*

1.4.2 Non-ribosomal peptide synthases

Non-ribosomal peptide synthases (NRPS) are, similarly to modular type I PKSs, modular megasynthases responsible for the production of secondary metabolites in bacteria and fungi. However, NRPSs produce metabolites by linking proteinogenic and non-proteinogenic amino acid units.

NRPS biosynthesis is mediated by a carrier protein-tethered thioester which activates the acyl group of each monomer unit and transfers the elongated chain between catalytic domains, whereas PKSs use malonyl-derived units to synthesize polyketide chains.

A minimal NRPS module consists of an adenylation domain (A), condensation domain (C) and a peptidyl-carrier protein (PCP)[72]. Firstly, the adenylation domain activates the carboxyl region of the amino acid with ATP, forming the acyl-phosphoric acid anhydride with AMP. This is followed by loading onto the phosphopantheteine moiety of the PCP. The condensation domain then catalyses the nucleophilic attack of the amino group of the previously activated amino acid to the carbonyl of the tethered acyl group from the previous module. This results in the formation of a new peptide bond between the two units (Figure 1.11.)[73]. Additional (optional) domains can be incorporated into a module for further modification to the intermediate. For example, a cylisation (Cy) domain can be used to produce thiazoline or oxazoline intermediates, which can be further modified by an oxidation (Ox) domain to yield the thiazole and oxazole products. Epimerization or E domains are among the most abundant modification domains in NRPSs and they catalyse the stereochemical inversion of L-amino acids to D-amino acids. (TE) domain either hydrolyses or cyclises the intermediate to yield the final product[74].

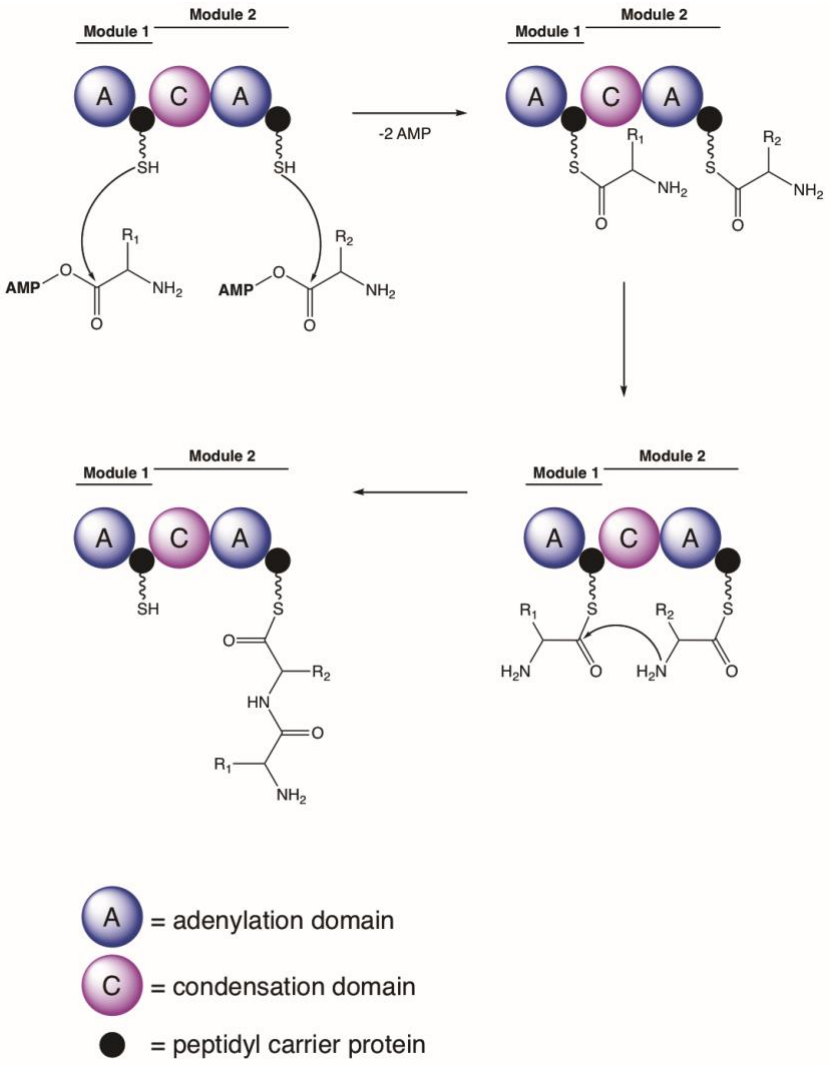


Figure 1.13 Schematic representation of NRPS domains mechanism.

1.4.2.3 Structure and functions of NRPS domains

Condensation domain

Condensation (C) domains represent a versatile class of NRPS domains. Any NRPS composed of more than one module contains at least one C-domain. However, single modular NRPSs may include C-domains, especially if they cooperate with other NRPSs. The primary target of a C-domain is the condensation of the upstream and downstream activated substrates through a nucleophilic attack, leading to the formation of a dipeptide and peptide chain elongation.

The C domain has two binding sites: one for the electrophilic donor substrate (the acyl group of the growing chain) and one for the nucleophilic acceptor substrate (the activated amino acid)[37]. Thus, it is accessible to substrates attached to the phosphopantetheine arms of both an upstream donor peptide carrier protein (PCP) and a downstream acceptor PCP. Also, several residues of the C-domain can fulfil multiple functions - condensation and epimerization, N-methylation, cyclization or serine-dehydration.

A channel running through the protein is formed at the interface of the two subdomains, starting near the floor loop of the C-terminal subdomain and ending near the helix α1 of the N-terminal subdomain with the conserved HHxxxD motif located at the midpoint of the channel (Figure 1.14)[75]. The PCP-C didomain structures reveal that the donor PCP is positioned on the C domain at the entrance near the floor loop[76,77]. The structure of termination module from *Acinetobacter baumannii* suggests that the acceptor PCP delivers the substrate into the substrate channel form the entrance near the helix α1 of the N-terminal subdomain[78]. The second histidine of the conserved HHxxxD motif is proposed to play an important role in positioning the α-amino group for nucleophilic attack[79].

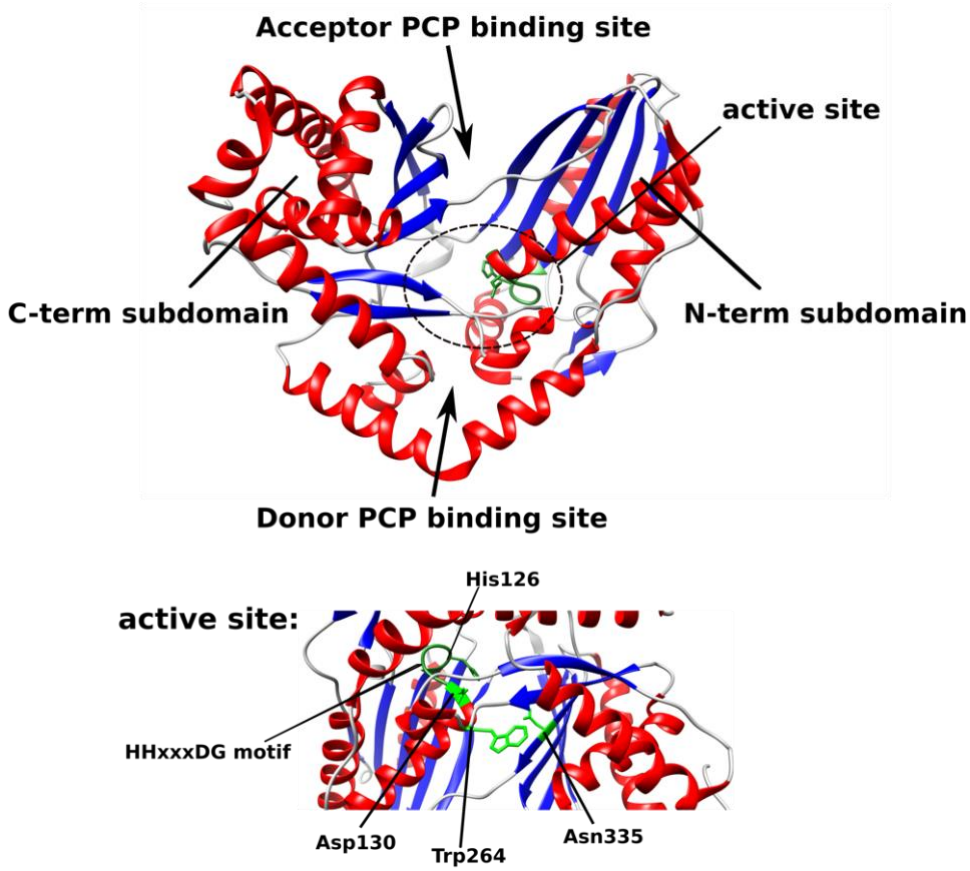


Figure 1.14 The structure of NRPS condensation (C) domain VibH[80]. The figure was created from the PDB file 1L5A using Chimera. The residues shown were identified by mutagenesis to be important for C-domain catalyitic activity. His 126 is not crucial for VibH catalysis. However, this residue is required in other C domains.

Peptidyl carrier proteins (PCPs)

The NMR solution structures of PCPs reveal a four-helix bundle with an extended loop between the first two helices, resembling the overall fold of acyl carrier proteins (ACPs) from PKSs (Figure 1.11). Also, the conserved serine, which is the cofactor-binding site, has the same location as in the ACPs (Figure 1.15). PCPs are sometimes referred to as thiolation domains. An important difference between ACPs and most PCPs is the sequestration of acyl chains within the core of ACPs (i.e. type II ACPs), which has not been observed for the substrates bound to PCPs[81,82]. ACP sequestration of substrates requires interacting domains to utilize a chain flipping mechanism to allow the bound substrates to contact these interacting domains, which is not required for PCPs during NRPS biosynthesis[83,84].

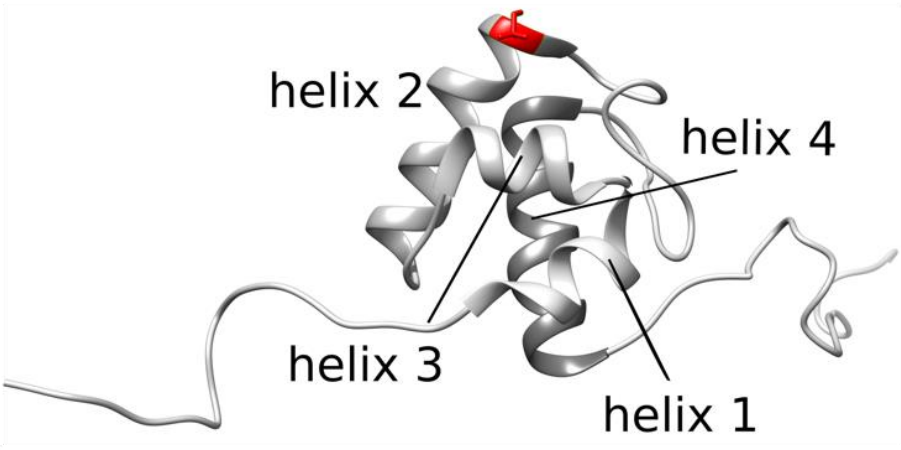


Figure 1.15 Structure of apo PCP1 from yersiniabactin synthetase (PDB: 5U3H). Conserved serine is highlighted in red.

1.4.3 Hybrid NRPS-PKS systems

The processive and modular biosynthetic logic of NRPSs and PKSs allows their integration into hybrid systems that are utilised by bacteria and fungi to synthesize natural products. Such systems are referred to as NRPS-PKS hybrids, and are responsible for the biosynthesis of many bioactive compounds, examples of which are shown in Figure 1.16.

1.4.4 Engineering of modular PKSs and NRPSs

The potential interchangeability of PKS and NRPS catalytic domains and direct communication between enzymes makes them an appealing target for combinatorial biosynthesis. Combinatorial biosynthesis generates novel molecules by mixing and matching parts of known natural products, manipulating the genes encoding the enzymes responsible for their biosynthesis and altering functional groups through the manipulation of biosynthetic enzymes. Thus, polyketides are now a major area of study due to the idea to use combinatorial biochemistry in drug discovery and development.

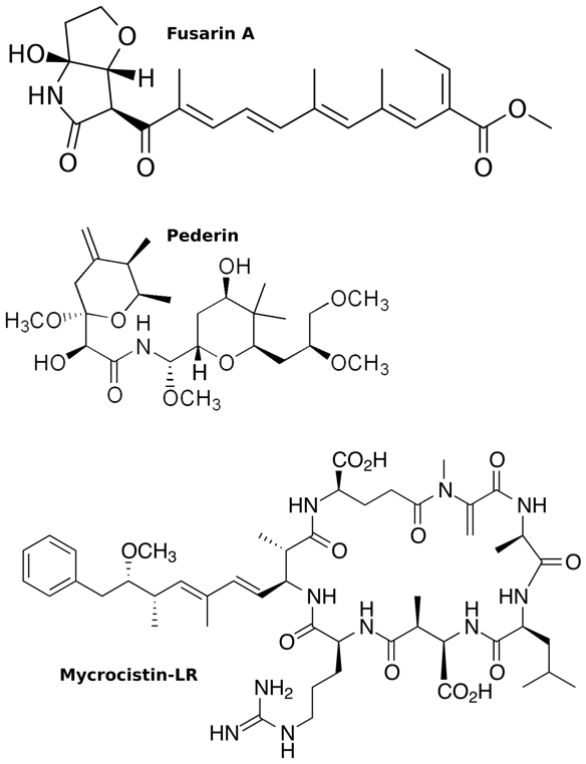


Figure 1.16 Natural products synthesised by NRPS-PKS systems. Pederin is an antiumor agent. Fusarin is a mycotoxin that infects agriculturally important crops. Mycrocistins are hepatotoxins.

One approach includes manipulation of individual PKS domains in active residues to adjust their substrate and stereochemical specificity. Secondly, switching one domain for another from the same or different PKS, or even altering entire modules might reduce or extend the elongating chain. Addition or removal of catalytic domains would affect substrate during elongation[85]. Domain exchange has already been obtained by swapping an AT in DEBS with an AT from the rapamycin PKS, which enabled the modification of all extender units incorporated into the erythromycin backbone[86]. However, numerous attempts to exploit PKSs for combinatorial biosynthesis have not been successful due to the lack of information about the structure, mechanism and structural dynamics of the interactions between PKS domains.

1.5 Enacyloxin biosynthesis

The proposed biosynthetic pathway for enacyloxin IIa is presented in Fig. 1.17[11]. It is produced by a hybrid of type I PKS and NRPS. The loading module in protein Bamb 5925 consists of a MT domain, a GNAT domain (malonyl transferase/de- carboxylase from the Gcn5 histone acyltransferase family) and an ACP domain. ACP initiates the biosynthesis with a propionyl starter moiety. The chain elongation continues in for ten rounds of chain extension from PKS modules. The enacyloxin PKS modules belong to *trans*-AT PKS as many of these modules do not include AT domains. The polyketide chain release is mediated by a non-elongating KS^0 within module 10. Module 10 belongs to the trans-AT PKS phylogeny. The enacyloxin intermediate is proposed to be transferred from the ACP of Bamb 5919, to the adjacent KS^0. The substrate transfer continues from the KS^0 domain to a downstream PCP domain. It is postulated that specific protein-protein interactions contribute to the directionality of this process. Chain release is accomplished by the condensation of the 3'-hydroxyl group of a shikimate derived dihydroxy-cyclohexane carboxylic acid (DHCCA) with the PCP-bound polyketide. The condensation reaction is catalysed by an NRPS-like C domain. The released intermediate then becomes the substrate for a series of tailoring enzymes (not shown in Figure 1.17)

It has been shown in Challis laboratory that the transfer of the polyketide intermediate from module 10 ACP to KS^0 is necessary for the C domain to catalyse polyketide chain release via esterification with DHCCA[87].

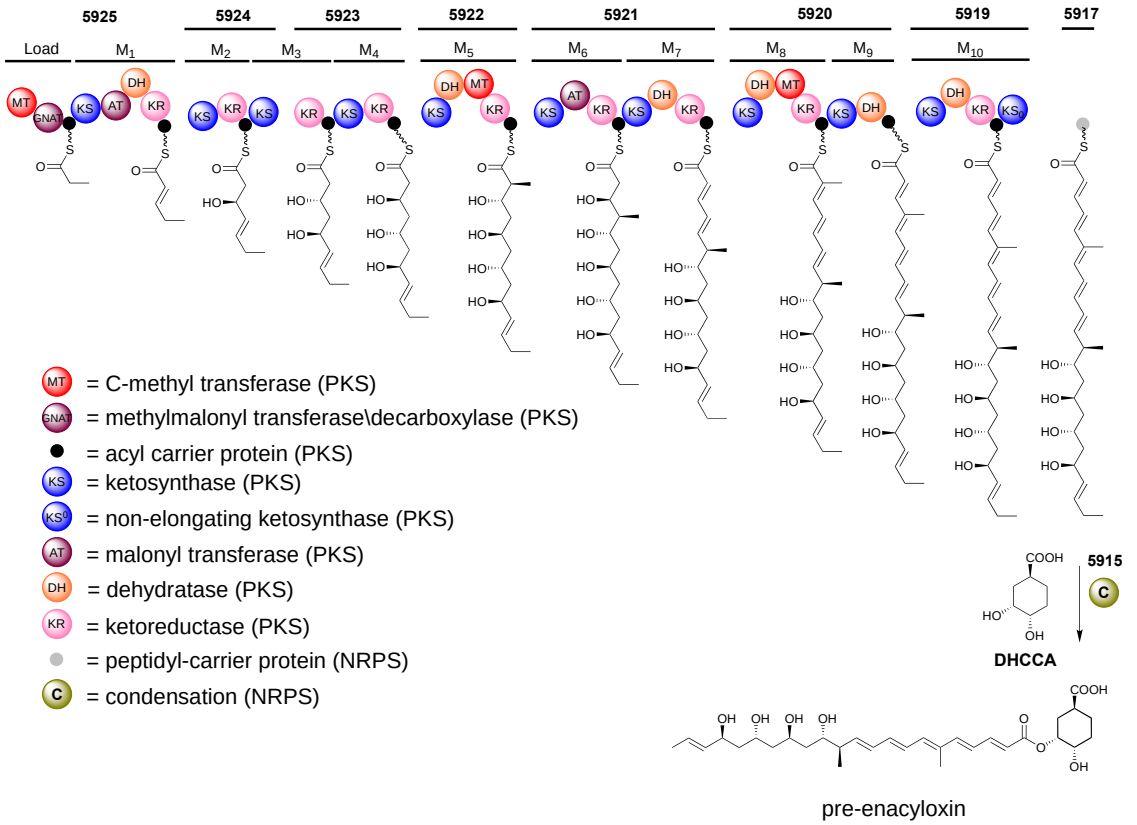


Figure 1.17 Proposed pathway for enacyloxin biosynthesis. The chain termination is mediated by the acyl carrier protein in module 10 (M10), which transfers the substrate to PCP via KS^0, which is also located in module 10. PCP then transfers the substrate to the condensation domain (C), which condenses the chain with DHCCA.

1.6 The aim of the project and objectives

As mentioned earlier in this chapter, enacyloxin IIa cannot be used clinically due to its structural instability. To create more stable and pharmaceutically useful enacyloxin IIa analogues, it might be possible to manipulate its biosynthesis by the approach based on mixing and matching components from different PKS and NRPS assembly lines. However, to harness enacyloxin IIa PKS-NRPS for that purpose, it is important to understand the molecular structures and dynamics responsible for the specificity of domain-domain interactions and directionality of biosynthesis. This is particularly important for the domains involved in controlling the crucial step of polyketide chain release where two separately assembled molecules are joined together through an ester bond. The ester bond introduced in this step is one of the sources of enacyloxin instability and ideally

should be replaced by a more stable functionality, e.g. amide bond, for pharmaceutical applications.

As mentioned in the previous section, the chain termination involves the unusual indirect transfer of the polyketide chain from the ACP in module 10 to the downstream PCP, which then delivers the substrate to the condensation domain using non-elongating KS^0, instead of direct interaction between ACP and condensation domain (C) (Figure 1.18). Since non-elongating KS^0 does not modify the polyketide chain, it acts only as a shuttle between module 10 ACP and PCP.

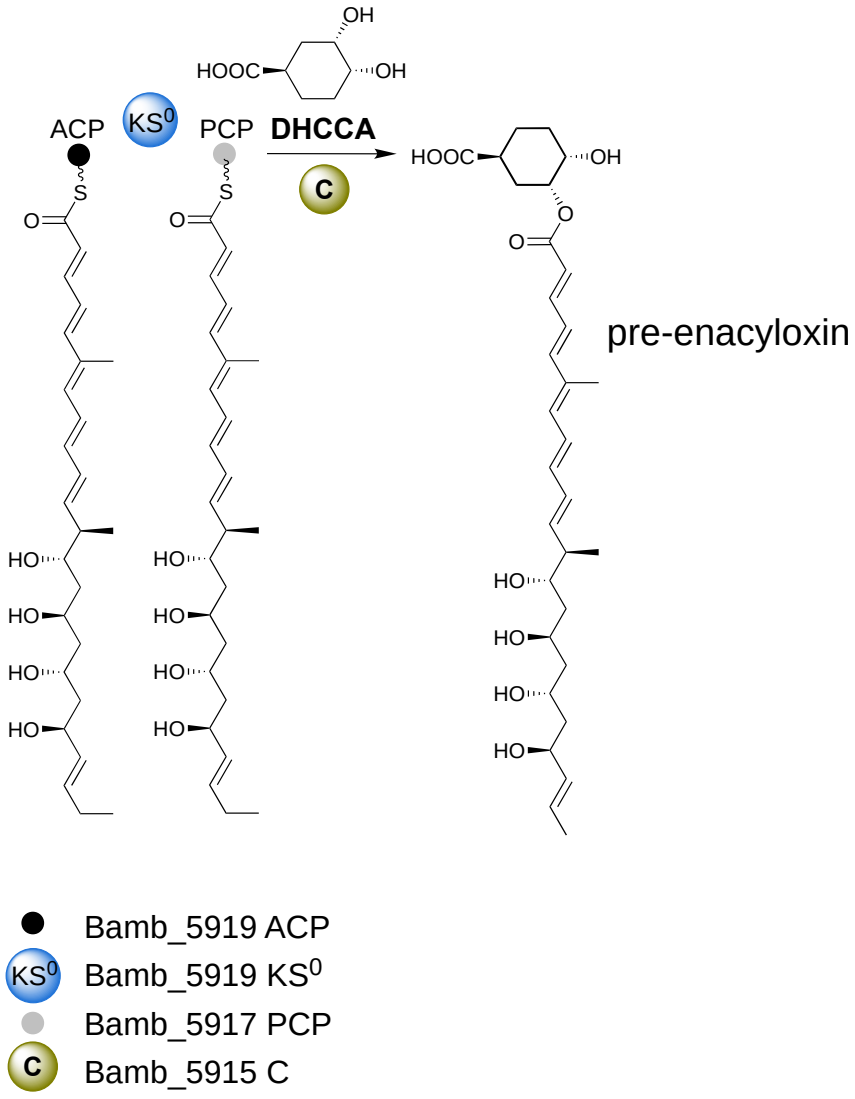


Figure 1.18 The chain termination in enacyloxin IIa biosynthesis (based on Figure 1.17).

As the chain termination process involves two surplus enzymes (KS^0 and PCP) which do not elongate or modify the substrate, it was proposed that a series of compatible interactions takes place between ACP, KS^0, PCP and C protein domains. Elucidating the molecular basis for product release from the enacyloxin pathway is thus important for the application of this unusual biosynthesis to synthetic biology approaches for novel antibiotic production. Generally, a major challenge for engineering biosynthetic pathways is developing a molecular-level understanding of domain-domain and protein-protein interactions to facilitate efficient catalysis. The description of interactions during enacyloxin chain release could explain rules of interactions that could enable bioengineering of novel biosynthetic pathways, and describe unidentified features from characterised biosynthetic pathways.

However, carrier proteins, consequently their complexes with KSs, are highly dynamic and challenging to crystallise and be studied by X-ray crystallography. On the other hand, NMR spectroscopy can provide the structural and dynamical insights required to understand which structural characteristics of enacyloxin PCP make it a preferable interacting partner with the condensation domain compared to ACP. Solution-state NMR can provide valuable information about domain-domain interactions. However, solution-state NMR analysis becomes substantially difficult for biomolecules over 30 kDa. This is not a problem for the analysis of enacyloxin carrier proteins, however, it poses a problem for the characterisation of their complexes with KS^0. There are no such size limitations placed for solid-state NMR. Hence, the overall aim of this project was the structural and dynamical characterization of the interactions of enacyloxin module 10 ACP and the downstream PCP with KS^0 by the combination of solution and solid-state NMR, followed up by biochemical and biophysical methods. The initial objective was to overproduce and purify ACP, KS^0 and PCP domains and confirm their bioactivity by mass spectrometry enzymatic assay. The second objective was to express isotopically labelled ACP domain to obtain NMR backbone assignments required to study its interaction with isotopically unlabelled KS^0 domain by solution-state NMR titrations. The third objective was to overproduce and purify isotopically labelled PCP domain and study its interaction interface with the KS^0 domain also by solution-state NMR titrations. The final objective was to investigate the interactions between ACP and PCP domains with KS^0 by using solid-state NMR spectroscopy.

Chapter 2. Materials and methods

2.1 Instruments and equipment

WPA CO8000 Cell density meter (Biochrom) was used to measure optical density of microbial cultures. NanoDrop Lite spectrophotometer (Thermo Scientific) was used to measure DNA and protein concentrations. A BioRad Power PAC300 was used for agarose gel electrophoresis. Mini-PROTEAN® Tetra Cell System (BioRad) was used for SDS-PAGE. UHPLC–ESI–Q-TOF–MS was performed using the MaXis II™ (Bruker) mass spectrometer. Circular dichroism spectra were acquired on a JASCO 1500 instrument. Solution-state NMR spectra were obtained using Avance II-700 (Bruker). Solid-state NMR spectra were obtained using 500 MHz Bruker Avance III, Bruker 600 MHz Bruker Avance II+ and 700 MHz Bruker Avance III HD spectrometers. ÄKTA Pure 15M (GE Healthcare Life Sciences) and ÄKTA Purifier 100 FPLC systems were used to purify proteins.

Other instrumentation includes New Brunswick Scientific Innova® 44 refrigerated incubator shaker, Astell compact top autoclave, Thermo Electron Corporation Heraeus® Function Line incubator, Eppendorf® Centrifuge 5804 R, Hitachi Koki model Himac CR22N High-Speed Refrigerated Centrifuge and Eppendorf™ 5424 Microcentrifuges.

2.2 Chemicals and reagents

General chemicals were purchased from Sigma-Aldrich, Thermo Fisher, Merck and VWR International. Ultrapure H_2O (milliQ®, Millipore) used for dilution was prepared by autoclaving at 121°C for 20 minutes.

2.3 Expression hosts

Table 2.1: *Employed E. Coli strains*

Strain	Manufacturer	Genotype	Description
One shot® TOP 10™	Invitrogen	F-mcrA Δ(*mrr-hsdRMS-mcrBC*) Φ80lacZ ΔM15 ΔlacX74 nupG	Plasmid propagation.
One shot® BL21 Star™ (DE3)	Invitrogen	F$^-$*omp*T *hsd*S_B (r_{B^-} ,m_B) *galdcmrne*131 (DE3)	Protein expression

2.4 Protein constructs

The constructs for the expression of domains from enacyloxin biosynthetic pathway were produced by Dr. Paulina Sydor. Bamb_5919 ACP (shorter construct) was designed by Dr. Simone Kosol.

2.5 Buffers and gels

All buffers for FPLC protein purifications were filtered using 500 mL Steritop-GP 45 mm bottle top filter (0.22 μM pore size, Merck Millipore). If required, antibiotics ampicillin (100μg/ml) and kanamycin (50 μg/ml) were added to media.

Table 2.3 *Buffers and gels used in this work*

1% agarose gel	1 g agarose, 10 mL 10X TBE buffer, 90 mL H_2O, 5 μL of GelRed™ Nucleic Acid Stain (Biotium)
10X Tris-borate-EDTA (TBE) electrophoresis running buffer	108 g Tris-HCl, 55 g boric acid, 20 mL 1 M EDTA
LB medium (1 L)	25 g Lysogeny Broth, Miller (Thermo Fischer)
LB agar (1 L)	25 g Lysogeny Broth, Miller (Thermo Fischer), 15 g Bacto agar (Sigma-Aldrich)
10xM9 salt solution (1 L)	64 g $Na_2HPO_4 \cdot 7H_2O$, 12 g KH_2PO_4, 5 g NaCl
M9 minimal media (1L)	100 mL 10X M9 salt solution, 2 mL 1M $MgSO_4$, 100 μL 1 M $CaCl_2$, 10 mL mL MEM vitamin solution 100X (Gibco®), 1000x micronutrient stock solution,

1000x micronutrient stock solution (50 mL)	ammonium molibdate ($3 \cdot 10^{-6}$ M), boric acid ($4 \cdot 10^{-4}$ M), cobalt chloride ($3 \cdot 10^{-5}$ M), copper sulphate ($1 \cdot 10^{-5}$ M), manganese chloride ($8 \cdot 10^{-5}$ M) and zinc chloride ($1 \cdot 10^{-5}$ M
Na^{2+} -affinity loading buffer	50 mM Tris-HCl, 300 mM NaCl (pH=8.0)
Na^{2+} -affinity elution buffer	50 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole (pH=8.0)
SEC purification buffer	20 mM Tris, 150 mm NaCl, pH=8.0
Analytical SEC purification buffer	50 mM NaPi, 100 mM NaCl, pH=7.0
Laemmli resolving gel (12%, 5 mL)	880 µL H_20, 2 mL 29:1 (30%) Acrylamide:Bisacrylamide, 1.04 mL Tris buffer (pH 8.8, 1.5 M), 50 µL 10% SDS, 50 µL 10% ammonium persulfate, 2 µL TEMED
Laemmli stacking gel (12%, 1 mL)	680 µL H20, 170 µL 29:1 (30%) Acrylamide:Bisacrylamide, 130 µL Tris buffer (pH 8.8, 1.5 M), 10 µL 10% SDS, 10 µL 10% ammonium persulfate, 1 µL TEMED
10X SDS-PAGE (protein gel) running buffer	30 g Tris-HCl, 144g Glycine, 10 g Sodium dodecyl sulfate, H_2O added up to 1L
6X SDS-PAGE loading dye	2 mL 1 M Tris-HCl pH=6.8, 2 mL glycerol, 0.4 mL β-mercaptoethanol, 0.02 g bromophernol blue, 0.4 g sodium dodecyl sulfate, 2 mL H_2O
NMR buffer	50 mM NaCl, 100 mM $NaH_2PO_4 \cdot 7H_2O$, 10% (v/v) D_2O, adjusted to pH=7.0

2.6 Molecular Biology

Preparation of chemically competent cells

All proteins were recombinantly overexpressed using One Shot® BL21 Star™ (DE3) strain of *E. coli*. Plasmids were produced in the One Shot® TOP 10™ strain. All strains of *E. coli* were produced in the lab from commercially available stocks (Table 2.1). 1 µL of frozen *E. coli* stock was used to inoculate LB (10 mL). 2 mL of the overnight culture (37°C, 180 rpm) was then used to inoculate 100 mL of LB and grown at 37°C, 180 rpm until the OD_{600} ~ 0.4 was reached. Cells were transferred into 50 mL Falcon tubes and left on ice for 20 minutes. They were recovered by centrifugation at 3000 rpm, 15 min, 4°C. The pellet was

resuspended in ice-cold 0.1M $CaCl_2$ buffer (10 mL per 1L of culture) and left on ice for 30 minutes. The cells were then centrifuged (3000 rpm, 15 minutes, 4°C). The pellet was resuspended in ice-cold buffer (0.1M $CaCl_2$ with 15% glycerol, 1 mL per 100 mL culture). After being left on ice for 5 minutes, 50 µL aliquots were flash frozen and stored at -80°C.

Transformation of chemically competent *E. coli*

Aliquots of competent cells were gently mixed with 1 µL of target plasmid-DNA (100-150 ng/µL) and incubated on ice for approximately 20 minutes. The heat-shock was initiated at 42 °C for a period of 45 seconds. After 5 minutes on ice, 200 µL of LB medium were added. Cells were incubated for 1 hours at 37 °C, 180 rpm. Different volumes of cell suspensions (10-100µL) were dispersed on LB agar plates with relevant antibiotic resistance for selection of transformants. Plates were incubated overnight at 37 °C.

Plasmid isolation

Resulting colonies were picked and used to inoculate 10 ml of LB. The cultures were incubated overnight (37°C, 180 rpm) and harvested by centrifugation (4000 rpm, 15 minutes). Plasmid isolation was then carried out using GeneJET Plasmid Miniprep Kit (Thermo Scientific). The concentration of the purified DNA solution was determined by a NanoDrop Lite spectrophotometer (Thermo Scientific) and stored at -20°C.

Agarose gel-electrophoresis

Purity of the plasmid DNA was assessed by agarose gel electrophoresis. 1% w/v agarose gels were prepared in TBE electrophoresis buffer by heating the mixture in a microwave oven. After the agarose had dissolved and the solution cooled, GelRed was added (5 µl per 100 ml). 2 µl of the DNA loading dye 6x (Thermo Scientific™) were added per 10µL of the sample before loading the gel (Table 2.3). The DNA was subjected to electrophoresis at 100 V for 60 minutes. Estimation of DNA length was done with the aid of 1 kbp DNA ladders (NEB). DNA bands were visualised under UV-light.

2.7 Protein expression, purification and characterization

Overproduction and purification of proteins in LB media

Protein expression always started with freshly transformed *E. coli* BL21 Star™ DE3 cells (Table 2.1) that were plated overnight (37°C). The colonies were used for inoculation of 5 mL pre-culture. The culture was grown overnight (37 °C, 180 rpm) and was used to inoculate 1L of LB which was grown to OD_{600} ~ 0.6 (4 to 5 hours, 37 °C, 180 rpm). Expression was induced by addition of 1 mM IPTG at an OD_{600} of 0.6-0.8. The incubation at 15-18 °C was proceeded overnight.

Expression of isotope labelled proteins

For the expression of ^{15}N and $^{13}C^{15}N$-labelled proteins, BL21 Star™ DE3 cells (Table 2.1) were first grown in 2 x 1 L of LB medium at 37 °C to an OD_{600} > 1.0. The cells were pelleted by centrifugation at 4,000 g for 15 minutes and washed with M9 salt solution (Table 2.3). Centrifugation step was repeated and the pellet resuspended in 20 mL of M9 medium supplemented with $^{15}NH_4Cl$ and [U-^{13}C]glucose (Sigma Aldrich). Resuspended cells were used to inoculate 1L of M9 medium. After incubation for 1 hour at 18 °C and 180 rpm, expression was induced with 1 mM IPTG, and the culture was incubated at 18 °C overnight.

Immobilised metal affinity chromatography protein purification

Overnight media cultures containing overexpressed proteins were centrifuged (7000 g, 20 minutes, 4°C). The resulting supernatant was loaded onto a HiTrap® Chelating Column (GE Healthcare), which had been pre-loaded with 100 mM $NiSO_4$ and equilibrated in loading buffer (Table 2.3). Column-bound proteins were eluted in 2 mL fractions using elution buffer linear gradient from 10 mM to 500 mM imidazole. Protein containing fractions were pooled and analysed on the SDS gel.

Size exclusion chromatography (SEC)

The final purification step involved separation of the proteins with respect to their molecular size by employing HiLoad® 16/600 Superdex® 200 pg (GE Healthcare) gel filtration column to purify Bamb_5919 KS^0 and HiLoad® 16/600 Superdex® 75 pg (GE Healthcare) column for the purification of Bamb_5919 ACP and Bamb_5917 PCP. The samples were loaded on *ÄKTA systems via* 1 mL sample loop (GE Healthcare). The elution with SEC buffer (1.5 CV) was performed at a flow rate of 1 mL/min (Fraction collector F9-C, GE Healthcare). The eluted volume was collected in 1.5 mL fractions, which were analyzed by SDS-PAGE.

Concentration and buffer exchange of protein samples by ultrafiltration

Protein containing fractions were concentrated with Amicon centrifugal filters (Millipore). The MWCO was selected depending on the size of the protein (MWCO < MW/2). Small sample volumes were concentrated in centrifugal Amicon Ultra-0.5 filter units (Millipore). Centrifugal filters were also used for buffer exchange and desalting.

Cleaving His-tags from proteins

Green fluorescent protein tobacco etch virus (GFP-TEV) was used to cleave the His-tag from purified proteins. Proteins and GFP-TEV were mixed in approximately 5:1 ratio and were incubated for 2 hours at 30°C to achieve the optimal TEV turnover rate[88]. TEV digested proteins were run through a Na^{2+}-affinity column or HiLoad® 16/600 Superdex® 75 pg (GE Healthcare) gel filtration column to remove remaining His-tagged proteins and GFP-TEV protease. The flow-through and fractions were collected and examined on an SDS-PAGE gel. Samples were then concentrated and stored. The plasmid containing TEV protease was obtained from dr. Muhammad Hasan.

Analytical SEC for protein characterization

Analytical Superdex® 200 Increase 10/300 GL column (GE Healthcare) was used to assess the molecular weight and the oligomeric state of Bamb_5919 KS^0 and Bamb_5917 PCP in

NMR buffer (Table 2.3). The column was equilibrated with 2 CV of NMR buffer (Table 2.3). Protein samples were injected via 2 mL sample loop (GE Healthcare) and eluted at a flow rate of 0.75 mL/min. The mixture of standard proteins (Sigma Aldrich) was applied to the columns for calibration (Table 2.4). The observed retention volumes were correlated with the known molecular weights of standard proteins. Blue Dextran (Sigma Aldrich) was used for the *determination* of the *void volume*. Elution time of each marker is displayed in Table 2.3.

Table 2.4 *Protein standards used for calibration*

Protein marker	Molecular weight (MW)	Elution volume (mL)
Thyroglobulin	660	9.09
Apoferritin	443	9.99
β-amilase	200	11.77
Bovine serum albumin (BSA)	66	13.84
Lysozyme	14	20.04

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

Polyacrylamide gel-electrophoresis (PAGE) was performed under denaturing conditions employing sodium dodecyl sulfate (SDS). Polyacrylamide percentage of 12% was used for analysis. Usually, 10 µL of protein samples and 25 µL of SDS-loading buffer (Table 2.3) were mixed and incubated at 95 °C for 5 min. 6 µL of sample was loaded into each lane of the stacking-gel.

Electrophoretic separation of proteins was accomplished in a Mini-PROTEAN® Tetra Cell System (BioRad) by applying a voltage of 180 V for 45 min in 1X protein running buffer (Table 2.3). Staining of gels was achieved by using InstantBlue Protein Stain (Expedeon). The used protein marker was PageRuler™ Prestained Protein Ladder (Thermo Scientific™). Gels were destained overnight in H_2O.

Storage of protein samples

100 µL sample aliquots were flash-frozen in liquid nitrogen and stored at -80 °C. Bamb_5919 ACP and Bamb_5917 PCP were stored in NMR buffer (Table 2.3) at the concentration of 500 µM. Bamb_5919 KS^0 and Bamb_5919 ACP_KS^0 were stored in NMR buffer with 10% glycerol at the concentration of maximum 50 µM to avoid protein precipitation.

Determination of protein concentration

The concentration of protein solutions was assessed by measuring the absorption of aromatic residues at a wavelength of 280 nm (A280nm). Drops of 3 µL were measured in triplicates using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). Extinction coefficients were obtained from the ExPASy ProtParam tool[89]
Bamb_5919 ACP with the His-tag removed does not contain tryptophans and tyrosines so the sample concentration could not be accurately determined by using A280nm absorption. Hence, these samples were exchanged first to water (2 mL final volume), flash-frozen and freeze-dried at –60 °C and 0.1 mbar for 8 h (Virtis® Brand Benchtop(TM) 'K' Freeze Dryer). Sample concentration was calculated based on the weighted protein mass. The dry protein powder was tightly sealed and stored at –80 °C.

2.8 Homology modelling of Bamb_5919_ACP

Homology model of ACP domain was constructed using the Phyre+ server[90]. The utilised template was the published structure 2JU2[91]. The model was visualized using UCSF Chimera package[92].

2.9 Biochemical assays

Phosphopantetheinylation of carrier proteins

Apo-Bamb_5919 ACP and *apo*-Bamb_5917 PCP were converted to their holo forms by using phosphopantetheinyl transferase (PPTase) Sfp from *B. subtilis*. 400 µM of each *apo*-protein was combined with 10 µM Sfp, 10 mM $MgCl_2$ and 1 mM CoA per 500 µL of the reaction solution. Reactions were carried out for an hour at a room temperature in an

NMR buffer at pH=7.0. Phosphopantetheinylation was verified by UHPLC-ESI-Q-TOF-MS analysis (maXis II Bruker). The plasmid containing Sfp was acquired from Dr. Manuela Tosin.

Acyl transfer assays

Apo-Bamb_5917 PCP was converted to its holo form as described above. *Apo*-Bamb_5919 was converted to its acetyl form by adding acetylCoA to the reaction mixture instead of CoA. Excess acetyl-CoA and CoA was removed by using Amicon Ultra 0.5 centrifugal filters (3 kDa). Acetyl transfer reaction was initiated by incubating 100 µM *holo*-Bamb_5919 ACP and *holo*-Bamb_5917 PCP with 25µM *holo*-Bamb_5919_KS^0. LC/MS experiments were performed using UHPLCESI-Q-TOF-MS analysis.

2.10 Analytical techniques

Circular dichroism (CD) spectroscopy

CD spectra were recorded between 190 and 260 nm at room temperature a JASCO J-1500 machine with a Peltier type cell holder (PTC-423L) using a 1-mm path length quartz cuvette, , with nine scans for data accumulation, a data pitch of 0.1 nm, band widths of 2 nm and a scanning speed of 100 nm/min. 500 µL (0.1 mg/mL) of protein samples were prepared in an NMR buffer (Table 2.2). The buffer was measured for baseline evaluation and correction.

Melting curves were recorded in a temperature range of 20 to 90 °C. The temperature scan was performed in increments of 0.1 °C and with a heating rate of 1 °C/min. CD spectra were recorded at the initial and final stages of the temperature scans as well as at the initial

temperature after the heating process to test for reversibility. Spectral analysis was performed using DichroWeb [92] using SELCON3 method for fitting[93].

Ultra-high performance liquid chromatography coupled with electrospray ionization-quadrupole-time of flight-mass spectrometry (UHPLC–ESI–Q-TOF–MS) protein analysis

Protein masses and acyl-transfer assays were analyzed on a MaXis II™ (Bruker) ESI–Q-TOF–MS connected to a Dionex 3000 RS UHPLC fitted with an ACE 3 C4-300 column (100 × 2.1 mm). The column was eluted with a linear gradient of 5–100% acetonitrile containing 0.1% formic acid over 30 min. The mass spectrometer was operated in positive ion mode with a scan range of 200–3,000 m/z. Compass DataAnalysis 4.1 (Bruker) was used to extract ion chromatograms.

Bio-layer Interferometry (BLI) measurements

The presence of only one lysine residue in the His6-tag was exploited for biotinylation of the Bamb_5919 ACP and Bamb_5919 PCP domain using the EZ-link Sulfo-NHS-Biotin cross-linking reagent (ThermoFisher Scientific), which reacts selectively with primary amines. After incubation for 30 min at room temperature, non-reacted cross-linking agent was removed on ZEBA spin desalting columns (3K MWCO, ThermoFisher Scientific) before loading Streptavidin biosensors (Pall ForteBio) with the biotinylated apo-Bamb_5919 ACP and apo-Bamb_5917 PCP domains at a concentration of 70 nM. The binding of Bamb_5919 KS^0 in concentrations of 0.175, 0.35, 0.75, 1.55, 3.1, 6.25, 12.5, and 25.0 µM to the immobilized biotinylated carrier protein domains was monitored on an Octet® RED96 System (Pall ForteBio) in triplicates, except for the binding of apo-Bamb_5919 ACP and Bamb_5919 KS^0. The sensor responses were fitted using the Octet Software. Affinity constants were extracted from steady-state fittings using the equation $\frac{B_{max}*X}{(X+K_d)}$, where X corresponds to the Bamb_5919 KS^0 concentration and B is the measured response in nm.

Isothermal titration calorimetry

ITC experiments were performed using a Microcal ITC200 (Malvern) located in the Biomedical Isothermal Titration Calorimetry Facility (Division of Virology at Addenbrooke's Hospital, University of Cambridge). The proteins were ran over a size-exclusion column first to exchange the samples to the NMR buffer. Bamb_5919 ACP and Bamb_5917 PCP were directly titrated into the chamber containing 200 µL of 25 µM Bamb_5919 KS^0. The concentrations of Bamb_5919 ACP and Bamb_5919 PCP were 12.5

μM, 6.25 μM, 3.1 μM, 1.55 μM, 0.75 μM, 0.35 μM, 0.175 μM. Nineteen 2 μL injections were performed with a pre-injection of 0.2 μL (total ligand volume of 38.2 μL, 240 s injection intervals). The stirring speed was 750 rpm. The experimental temperature was 25°C. The data were baseline corrected and integrated with NITPIC, fitted with a nonlinear least squares routine using a single-site binding model[94].

2.11 NMR spectroscopy

2.12 Solution-state NMR spectroscopy

Solution NMR data were recorded on a 700 MHz Avance II (Bruker) spectrometer equipped with a cryogenically cooled probehead. Data acquisition and processing was done using TopSpin 3.5 (Bruker Biospin). Further data analysis and resonance assignments were performed using the software Cara[95]. All the samples were locked against the deuterium signal (10% (v/v) D_2O) and contained 1 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) for internal referencing. All solution-state NMR experiments were performed in a buffer containing 50 mM NaCl, 100 mM NaPi, pH = 7.0.

1D experiments

1D proton spectra were acquired using WATERGATE[11] for water suppression. Spectra were acquired using 16k complex data points and 64 co-added scans. The data were zero-filled to 32k and processed using a squared sine window function.

Backbone Resonance Assignment of apo-Bamb_5919 ACP

U-^{13}C-,^{15}N-labelled Bamb_5919 ACP was dissolved in 500 µL of NMR buffer (Table 2.3) to obtain a protein concentration of ~700 µM and pippeted into 3 mm NMR tubes (Wilmad®, Sigma Aldrich). Table 2.5 summarizes the NMR experiments and relevant acquisition parameters.

Table. 2.5 NMR experiments performed for apo-[U-^{13}C,^{15}N] Bamb_5919_ACP

Experiments	FID	NS	Spectral width (ppm)	Miscellaneous
^{1}H-^{15}N HMQC[96]	2048x128	16	35 (^{15}N)	
HNCO[97]	1114x64x120	16	35 (^{15}N) x 20 (^{13}C)	
CBCA(CO)NH[98]	2048x56x116	48	35 (^{15}N) x 80 (^{13}C)	
HNCA[99–102]	2018x48x96	16	35(^{15}N) x 30(^{13}C)	
HNCACB[103]	2048x64x128	48	35(^{15}N) x 80(^{13}C)	Non-uniform sampling
(H)CCH-TOCSY[104]	2048x56x200	16	80(^{15}N) x 80(^{13}C)	
NOESY-HSQC[101,105,106]	2048x48x160	16	40(^{15}N) x 16(^{13}C)	100 ms mixing time

Solution-state NMR titrations

For protein-protein interaction studies, ^{15}N-labelled Bamb_5919 ACP (long construct), Bamb_5919 ACP and Bamb_5917 PCP were dissolved to a concentration of 100 µM. Bamb_5919 KS^{0} (350-450 µM) was added at each titration step to obtain the concentrations of 9.79, 19.19, 28.21, 45.21, 64.70, 82.50, 98.81, 113.81, 127.64 and 140.45 µM. A 2D ^{1}H-^{15}N SOFAST-HMQC[96] spectrum was acquired at each titration point. The affinity (K_d) of each residue was calculated individually based on the decay in integral value[107] according to:

$$[P] = \frac{A\left(-K_d - [L]_t + [P]_t + \sqrt{4K_d[P]_t + (K_d + [L]_t - [P]_t)^2}\right)}{2[P]_t},$$

where P defines dissociation of the carrier protein from its complex with Bamb_5919 KS^{0} at its single site (L). *A* stands for a multiplier constant that takes into account error in the integral of the first data point. The fitting was performed in Matlab R2107b by using the script provided by prof. J.R.Lewandowski.

To define the likely interacting residues, the mean K_a value ($K_a=\frac{1}{K_d}$) was calculated for all the residues and those with a K_a value larger than the mean + *3*σ were considered outliers. After no residues with K_a values larger than the average plus 3 s.d. were left, the mean was recalculated *and* the threshold defined as the mean + σ. Errors were propagated using Monte Carlo error analysis.

^{15}N relaxation measurements

U-^{15}N-labelled Bamb_5919_ACP was dissolved in 500 µL of NMR buffer (Table 2.3) to obtain a protein concentration of ~700 µM.

The ^{15}N longitudinal (T_1) values were measured using pseudo-3D pulse sequence *hsqct1etf3gpsi3d* and delay times were set at 19.96, 33.92, 50.88, 67.84, 84.80, 101.76, 135.68, 169.70, 203.52, 237.44, 305.28 and 339.92 ms. The ^{15}N transversal (T_2) values were measured using pseudo-3D pulse sequence *hsqct2etf3gpsi3d*, with the T_2 relaxation delays corresponding to 0.035, 0.075, 0.125, 0.3, 0.4, 0.5, 0.8, 1.5, 2 and 3 s. Other relevant experimental parameters are listed in Table 2.6.

Table 2.6. Parameters used in NMR ^{15}N dynamics measurements

Experiments	FID	Number of scans	Spectral width (ppm)
^{15}N R_1	4096x128	16	14 (^{1}H) x 38(^{15}N)
^{15}N R_2	4096x128	16	14 (^{1}H) x 38(^{15}N)
^{15}N–^{1}H NOEs	4096x512	64	14 (^{1}H) x 35(^{15}N)

Relaxation rates for each residue were derived from fitting curves with mono-exponential decays using Levenberg Marquardt algorithm in OriginPro 2018.

The {^{1}H}-^{15}N heteronuclear NOE was measured by employing the pulse sequence hsqcnoef3gpsi, and the ^{15}N-^{1}H NOE values were obtained by the ratio of peak intensities with and without a proton saturation.

2.13 Magic-angle-spinning (MAS) NMR spectroscopy

Analysed protein complexes were prepared by mixing 100 µM U-^{15}N-labelled Bamb_5919 KS^0 and 200 - 500 µM carrier protein in 500-700 µL NMR buffer volume (Table 2.3, pH=6.5). The samples containing gadolinium diethylenetriaminepentaacetic acid bismethylamide (Gd(DTPA-BMA)) were preparing by adding 0.2 mM Gd(DTPA-BMA).

Samples were incubated for 30 min at room temperature and then centrifuged (12000 g, 10 minutes, 4°C) to remove impurities. The sedimentation followed at 300000 g - 700000

g and 4°C for 24-120 hours (Optima™ Max-XP Tabletop Ultracentrifuge and MLA-150 Fixed-Angle rotor, Beckmann). Resulting supernatant was removed and used to dissolve 15 mg of DSS. 50 µL was carefully pipetted back into the centrifugation tube to prevent dehydration of the sediment and to introduce DSS into the sample during rotor packing. Sediments were manually packed into 0.8 mm or 1.3 mm rotors by using dedicated packing tools.

Solid-state NMR experiments were recorded at 500 MHz Bruker Avance III and 700 MHz Bruker Avance III HD spectrometers using a 1.3 mm Bruker probe in HCN configuration at 60 kHz MAS. 100 kHz MAS experiments were recorded at Bruker 600 MHz Bruker Avance II+ spectrometer using a 0.80 mm double-resonance probe from Samoson laboratory. A Bruker BCU-X cooling unit was used to regulate the internal sample temperature to 28-35 °C (as indicated from the chemical shift of water with respect to DSS). The magic angle and the homogeneity of the magnetic field were optimized by using alanine. Topspin 3.5 was used to process NMR data.

For the ^{1}H-^{15}N cross-polarization (CP) experiment carried out at 60 kHz, 90° pulses were set to 2.5 µs at 100 kHz nutation frequency for ^{1}H and 3.0 µs at 83.3 kHz nutation frequency for ^{15}N. At 100 kHz, the hard pulses were set to 2.0 µs at 125 kHz nutation frequency for ^{1}H and 3.0 µs at 83.3 kHz nutation frequency for ^{15}N. WALTZ-16 decoupling at 10 kHz was applied on protons during ^{15}N evolution and on ^{15}N channel during direct ^{1}H acquisition, while quadrature detection was achieved using the States-TPPI method. Suppression of the water signal was achieved by saturation 120 ms of MISSISSIPPI at an amplitude of half the MAS frequency on resonance with the water signal[108].

Chapter 3 - Interaction of acyl carrier protein and non-elongating ketosynthase from module 10 in enacyloxin PKS

3.1 Introduction

As discussed in Chapter 1, a key step in enacyloxin biosynthesis is the transfer of polyketide chain from an ACP domain in module 10 to a standalone peptidyl carrier protein by a KS^0 domain, which is necessary for the condensation domain to catalyse polyketide chain release via intermolecular esterification with DHCCA. Thus, the KS^0 transfers the substrate from a carrier protein (ACP), which is unable to interact with the C domain, to the carrier protein that can establish the interaction (PCP).

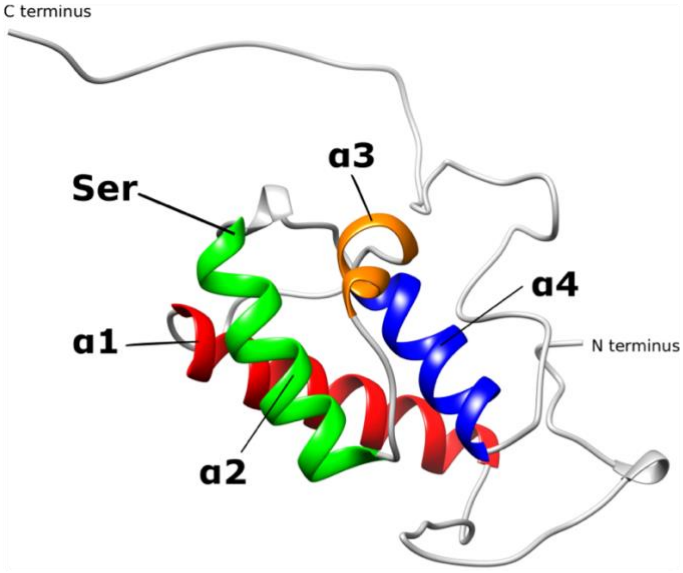
The aim of this chapter was to analyse the interaction interface between ACP and KS^0 as the first step towards the characterisation of the overall ACP-KS^0-PCP interface, i.e understanding what structural aspects make PCP a favourable interaction partner with C domain compared to ACP. This comparison is presented in Chapter 5.

Firstly, the NMR backbone assignment of *apo*-ACP was required to provide a 'map' of the protein which allowed the effects of KS^0 binding on ACP to be investigated directly.

Previously, the ACP protein domain construct was created by Dr. P. Sydor in the Challis lab. The construct comprised a sequence beginning at residue 1614 and ending at residue 1800 of the original endogenous protein domain (Supplementary Figure S.1). This construct was used in the work mentioned in Chapter 1 (section 1.5), where it was shown that ACP does not establish contact with C domain[87]. However, to facilitate solution-state NMR structural analysis, a shorter construct was created by Dr. S. Kosol (beginning at 1626 and ending at 1769 of the native domain) with truncated unstructured tails which were supposed not to hinder the interaction with KS^0 (Supplementary Figure S.1). The construct created by Dr. Sydor is referred to in the text as 'long Bamb_5919 ACP' and the construct created by Dr. Kosol as 'short Bamb_5917 PCP. The KS^0 construct, termed 'Bamb_5919 KS^0'in the text, was also provided by Dr. Sydor.

The homology model of short apo-Bamb_5919 ACP created by Phyre[2] [109] server revealed a typical four-helical protein fold characteristic of carrier proteins (Figure 3.1).

A



B

```
                                                Helix 1
ACP19|      GAPRVERRAGGTAGAALLAGLASLDAAARAARLKRHLEAAIRKLL-NRADTLDDRASMFD        59
2KR5|       ---------------AMAKG--VGVSNEKLDAVMRVVSEESGI----ALEELTDDSNFAD        39
2AF8|       -------------------MAT---LLTTDDLRRALVECAGETDGTDLSGDFLDLRFED        37
2PNG|       ------------------------GDGEAQRDLVKAVAHILGI---RDLAGINLDSSLAD        33
1KLP|       ----------------------MPVTQEEIIAGIAEIIEEVTGI---EPSEITPEKSFVDD        36
2DNW|       ------------GSSGSSGMPPLTLEGIQDRVLYVLKLYDKI---DPEKLSVNSHFMKD        44
2FVE|       ----------------------AKKETIDKVSDIVKEKLAL---GADVVVTADSEFSK        33
2FQ0|       ------------------------LKSTFDDIKKIISKQLSV---EEDKIQMNSNFTKD        32
1HY8|       -------------------------ADTLERVTKIIVDRLGV---DEADVKLEASFKED        31
2N57|       ----------------------MSDTAERVKKIVVEHLGV---DADKVTEGASFIDD        32
4IHF|       ----------------------SHMSTIEERVKKIIGEQLGV---KQEEVTNNASFVED        34
                                                 :  :                .
                      Helix 2          Helix 3          Helix 4
ACP19|      LGLDSLLSIDLRMQLEKDLACSLSTTVLHDHP--TIEALAGFLAERV----GAPPAGTVR        113
2KR5|       MGIDSLSSMVIGSRFREDLGLDLGPEFSLFIDCTTVRALKDFMLGSGDAG----------        89
2AF8|       IGYDSLALMETAARLESRYGVSIPDDVAGRVD--TPRELLDLINGALAEAA---------        86
2PNG|       LGLDSLMGVEVRQILEREHDLVLPIREVRQLTLRKLQEMSSKAGSD--TELAAPKSKN--        89
1KLP|       LDIDSLSMVEIAVQTEDKYGVKIPDEDLAGLR--TVGDVVAYIQKLEENPEAAQALRAK        94
2DNW|       LGLDSLDQVEIIMAMEDEFGFEIPDIDAEKLM--CPQEIVDYIADKKDVYESGPSSG---        99
2FVE|       LGADSLDTVEIVMNLEEEFGINVDEDKAQDIS--TIQQAADVIEGLLEKKA---------        82
2FQ0|       LGADSLDLVELIMALEEKFNVTISDQDALKIN--TVQDAIDYIEKNNKQ-----------        79
1HY8|       LGADSLDVVELVMELEDEFDMEISDEDAEKIA--TVGDAVNYIQNQQ------------        76
2N57|       LGADSLDTVELVMAFEEEFGVEIPDDAAETIL--TVGDAVKFIDKASA------------        78
4IHF|       LGADSLDTVELVMALEEEFDTEIPDEEAEKIT--TVQAAIDYINGHQA------------        80
            :. ***  :      .      :

ACP19|      AGAAGGAGAGTGAPAGATGAAAAHAVSSASP     144
2KR5|       -------------------------------     89
2AF8|       -------------------------------     86
2PNG|       -------------------------------     89
1KLP|       IESE--------NPDAVANVQARLEAES--K     115
2DNW|       -------------------------------     99
2FVE|       -------------------------------     82
2FQ0|       -------------------------------     79
1HY8|       -------------------------------     76
2N57|       -------------------------------     78
4IHF|       -------------------------------     80
```

C

```
ACP19_long      ---------------GAPRVERRAGGTAGAALLAGLASLDAAARAARLKRHLEAAIRKLL   45
ACP19_short     LAELVEAEAGATGASGAPRVERRAGGTAGAALLAGLASLDAAARAARLKRHLEAAIRKLL   60
                                ********************************************

ACP19_long      NRADTLDDRASMFDLGLDSLLSIDLRMQLEKDLACSLSTTVLHDHPTIEALAGFLAERVG   105
ACP19_short     NRADTLDDRASMFDLGLDSLLSIDLRMQLEKDLACSLSTTVLHDHPTIEALAGFLAERVG   120
                ************************************************************

ACP19_long      APPAGTVRAGAAGGAGAGTGAPAGATGAAAAHAVSSASP----------------------   144
ACP19_short     APPAGTVRAGAAGGAGAGTGAPAGATGAAAAHAVSSASPVPAGAASAAASAASAAAAAGA   180
                ************************************

ACP19_long      -------         144
ACP19_short     PSRATFA         187
```

Figure 3.1 Homology model of apo-Bamb_5919 ACP (A). Position of conserved serine residue is indicated. 98 residues (68% sequence coverage) have been modelled with 99.9% confidence by the single highest scoring template.
Multiple sequence alignment of ACPs with the known PDB structure from different systems (B). ACP19 – short Bamb_5919 ACP sequence. PDB identifiers: 2KR5 - Aspergillus parasiticus, 2AF8 - Streptomyces coelicolor, 2PNG - Rattus norvegicus, 1KLP - Mycobacterium tuberculosis, 2DNW - Homo sapiens, 2FVE - Spinacia oleracea, 2FQ0 - Plasmodium falciparum, 1HY8 - Bacillus subtilis, 2N57 - Brucella melitensis, 4IHF - Escherichia coli.
The alignment of the long and short Bamb_5919 ACP construct reveals that the long construct contains longer terminal tails(C). The regions predicted as helical are highlighted. The alignments were done in ClustalΩ[110].

3.2 Biophysical and biochemical characterization of Bamb_5919 ACP and Bamb_5919 KS^0

3.2.1 Protein expression and purification

All constructs were expressed with N-terminal His affinity tag of 34 residues to facilitate protein purification. The tag sequence comprised the Tobacco Etch Virus protease (TEV) cleavage site EXXYXQ-S/G (TEVp cleavage Site, TS) where X can be any amino acid[111]. The cleaved proteins also contained residues GIDPFT belonging to His-tag. Bamb_5919 ACP and Bamb_5917 constructs that were used for NMR studies always had the His-tag removed. The expected masses of protein constructs are listed in Supplementary Fig. S.2. The purification of Bamb_5919 PCP is described in this section since it was used for the bioactivity assay described in this section. As described in Chapter 2, the supernatant resulting from cell-free extracts was firstly purified on a Ni^+ affinity column (Figure 3.2.C). The tag was then cleaved (Figure 3.2.A, D) and fractions further purified by size-exclusion chromatography to ensure high purity (Figure 3.2 B, D, E). 11 mg of purified Bamb_5919 ACP was obtained from 1 L of LB and 8 mg was obtained from M9 *E. Coli* cultures (^{15}N and $^{15}N^{13}C$-labelled protein), as determined by dry weight (after liophylisation) of the purified protein pellet. The same yields were obtained for the long construct. Contrary to the long construct, short Bamb_5919_ACP did not retain any tyrosines or tryptophans in its protein sequence after His-tag removal so it was not possible to use A_{280} absorption as a reliable protein concentration measurement reference. Bradford assay was attempted, but the obtained measurements were not consistent, probably due to the

quality of the reagent and pipetting errors. The concentrations of proteins with high enough extinction coefficients (Bamb_5919 KS^0 and Bamb_5917 PCP) that were calculated based on A_{280} absorption were the same as those calculated from the dry weight of the proteins. For that reason, this method was used to estimate the mass and concentration of Bamb_5919 ACP. Bamb_5919 ACP appeared to be stable following sequential freeze-thawing at room temperature for 30 days as seen by the lack of precipitation and subsequently recorded ^{1}H-^{15}N-HMQC spectra.

The Bamb_5919 KS^0 domain was also expressed with an N-terminal tag for facilitated purification on Ni+ affinity column. To ensure higher purity, it was further run through the size exclusion column (Figure 3.2.E). The tag was not cleaved as the total contribution of 34 His-tag amino acids was not supposed to significantly interfere with the overall protein (670 amino acids) behaviour. The total yield was 30 mg from 1L of LB medium. To avoid precipitation, protein stocks were kept at the concentration of 50 μM and concentrated to a maximum of 350 μM if needed, usually for NMR titration experiments. Bamb_5917 PCP was purified using Na^{2+}-affinity and gel-exclusion chromatography (Figure 4.2.), after His-tag removal. Approximately 13 mg of protein was obtained from 1 L of LB culture and 11 mg from M9 culture. Freeze-dried Bamb_5917 PCP appeared to be stable following sequential freeze-thawing and was not subject to proteolysis.

LC-MS spectra of the purified proteins are attached in Supplementary S.2.

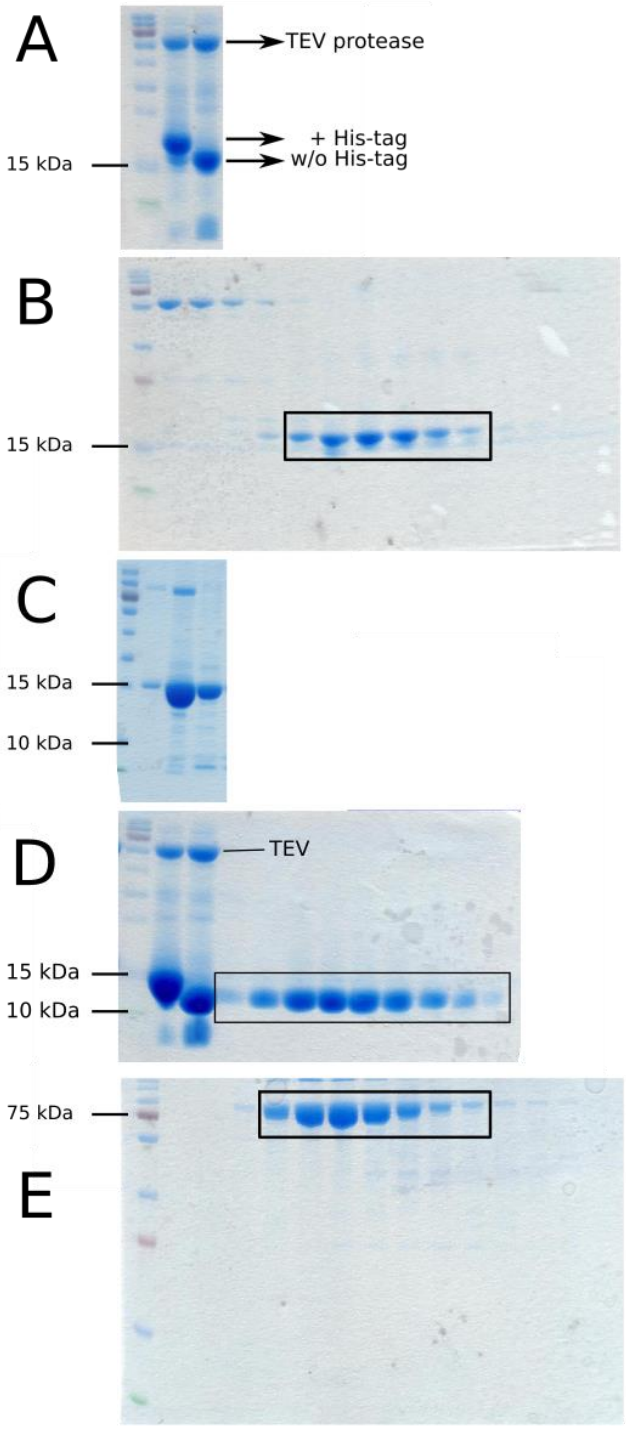


Figure 3.2 12% SDS-PAGE gels showing the purification and His-tag cleavage of the proteins used in the project. A) His-tag cleavage of short Bamb_5919 ACP. B) The samples taken from each eluted fraction after the size-exclusion purification of short-Bamb_5919 ACP with a removed His-tag. C) Long Bamb_5919 ACP fractions after Ni^{2+} affinity purification. D) The samples of Bamb_5917 PCP taken after gel-exclusion chromatography. The gel also shows the cleavage of a His-tag. E) Bamb_5919 KS^{0} purified after size-exclusion chromatography.

3.2.2 Analysing thermal stability of Bamb_5919 ACP and Bamb_5119 KS^0 by circular dichroism

As NMR analysis was the main experimental tool used in the project, it was important to understand the effect of temperature on the integrity of the samples, i.e. secondary structure and fold. NMR experiments are usually run for several days so it was important to ensure that the samples would not degrade or have their secondary structure altered if exposed to the targeted experimental temperature.

Circular dichroism (CD) is an excellent technique for rapid characterisation of the secondary structure and folding properties of purified proteins and is used to determine whether the purified protein is folded. CD can also be used to monitor conformational changes due to temperature[112]. However, CD does not give the residue-specific information that can be obtained by X-ray crystallography or NMR, but provides a more robust way to assess temperature effects on protein fold.

Protein secondary structure can be determined by CD in the spectral region between 190 nm and 250 nm, which is within UV region of the electromagnetic spectrum. The main chromophore in this region is the amide linkage of peptide bond[113]. α-Helix, β-sheet, and random coil structures give rise to a characteristic shape of CD spectrum. However, CD spectra provide the global picture of the protein secondary structure and it is not possible to determine the position of a backbone amino acid within a protein. Therefore, the CD spectra are generally used to assess the effect of different experimental conditions on proteins.

CD analysis of apo-Bamb_5919 ACP (Figure 3.3A) revealed that an increase in temperature from 20 °C to 50 °C did not significantly affect its fold. Apo-Bamb_5919 ACP seems to be mostly helical (~40%) at 20 °C to 40 °C (Figure 3.1A). It was not possible to determine the melting temperature from the melting curve (Figure 3.1B). The CD analysis of Bamb_5919 ACP was performed on apo protein, but the addition of Ppant arm should not significantly affect the protein fold. It thus was assumed that secondary structure characteristics of holo Bamb_5919 ACP should be similar to apo-Bamb_5919 ACP across the same temperature range.

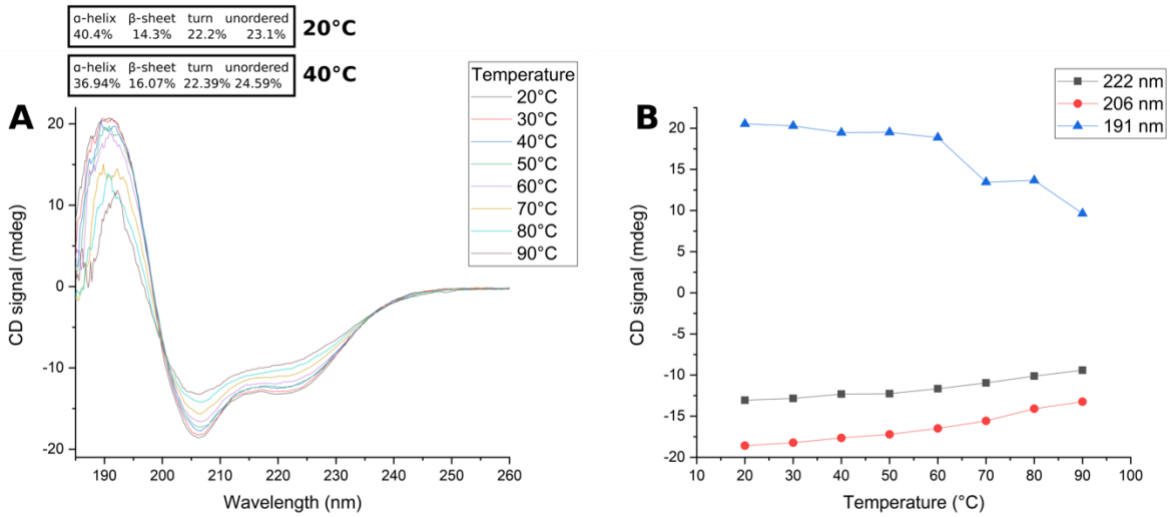


Figure 3.3 CD analysis of apo-Bamb_5919 ACP. A) CD spectra recorded at a temperature range between 20 °C and 90°C. Secondary structure analysis (Dichroweb) at 20°C and 40°C is indicated. B) Melting curves at 222 nm, 206 nm and 191 nm.

On the other hand, the secondary structure of Bamb_5919 KS^0 seems to be more susceptible to changes when temperature increases to 40 °C, compared to Bamb_5919 ACP. CD spectra recorded at 40°C (Figure 3.4A) showed that the secondary structure profile is significantly different compared to its profile at 20°C. The melting temperature estimated from the melting curve at 206 nm was between 50 and 60°C.

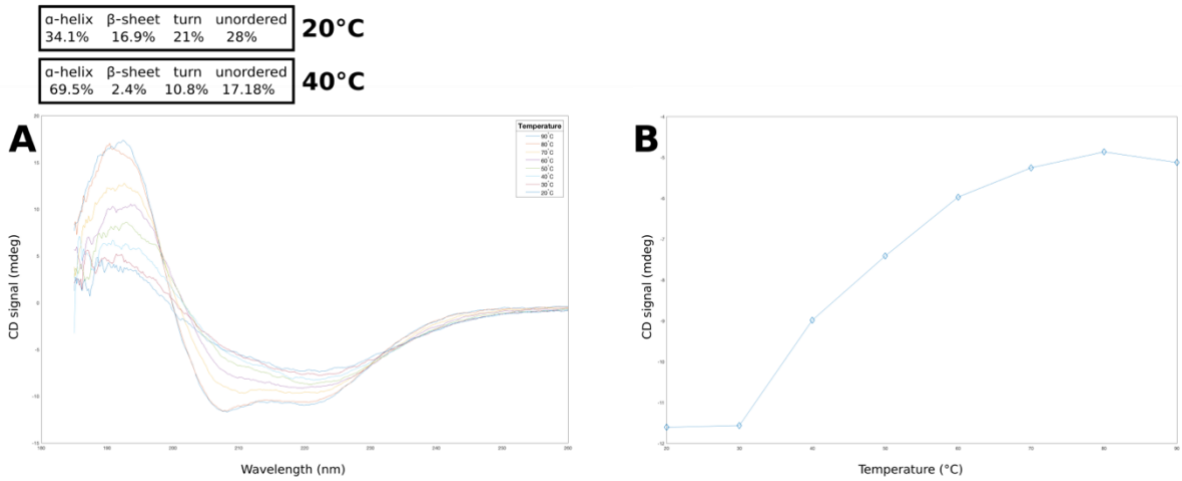


Figure 3.4 CD analysis of apo-Bamb_5919 KS^0. A) CD spectra recorded at a temperature range between 20 °C and 90°C. Secondary structure analysis (Dichroweb) at 20°C and 40°C is also shown. B) Melting curve at 206 nm.

Thus, CD indicated that increasing the temperature during NMR measurement above 30℃ might affect the fold of Bamb_5919 KS^0 and potentially its function.

3.2.3 Characterisation of long and short Bamb_5919 ACP

It has been mentioned in the introduction of this Chapter that the short Bamb_5919 ACP construct was created to facilitate NMR analysis. However, it was important to ensure that this constructs retained its bioactivity after the terminal parts of the long Bamb_5919 ACP had been removed. For that purpose, solution-state NMR titrations were used. It was initially planned to titrate both long and short ^{15}N-Bamb_5919 ACP with Bamb_5919 KS^0 and monitor whether the excised parts of the tails might be involved in the interaction. The relevant information related to the analysis of protein-protein interactions by solution-state NMR titrations is provided later in the Chapter.

Firstly, to inspect whether the constructs were properly folded, ^{15}N-HMQC spectra were taken. The theoretical details about HMQC and HSQC spectra is provided later in the chapter. Compared to short Bamb_5919, long Bamb_5919 ACP produced a greater number of overlapped signals clustered in the central region of the spectrum (^{1}H: 8 – 8.5 ppm) (Figure 3.5). This was due to the presence of long unstructured terminal parts of long Bamb_5919 ACP. Generally, the residues located in that region are likely to be disordered. Unfortunately, due to that signal overlap, it was not possible to extract intensities of the individual peaks located in the central part of the spectrum and conclude whether termini in long Bamb_5919 ACP are crucial for the interaction with Bamb_5919 KS^0.

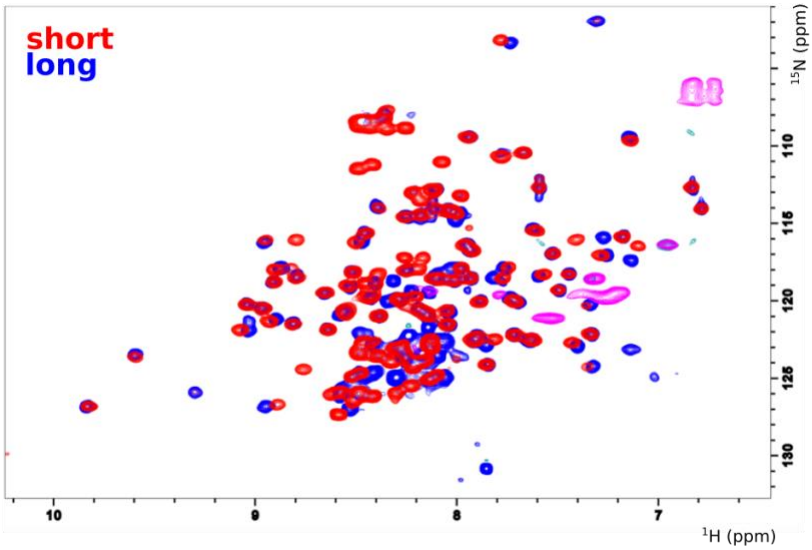


Figure 3.5 ^{1}H-^{15}N-HMQCs of the long and short apo-Bamb_5919 ACP construct. The spectra were recorded at 700 MHz and 298 K in the buffer containing 50 mM sodium phosphate, 50 mM NaCl and 10 % D_2O.

3.2.4 Bioactivity assay

The NMR analysis of the long and short Bamb_5919 ACP discussed in the previous chapter did not provide enough information to determine whether the short Bamb_5919 ACP is functional. Hence, to confirm its bioactivity and to demonstrate that the truncated disordered tails are not essential for function, the transfer of a substrate from holo-Bamb_5919 ACP to holo-Bamb_5917 PCP was reconstituted in vitro in the presence and absence of Bamb_5919 KS^0. The bioactivity of the His-tagged long Bamb_5919 ACP had already been confirmed by dr. P. Sydor[87] by the same assay. This reaction was thus used as a positive control. Also, all analysed domains contained His-tag. Firstly, carrier proteins were converted from apo to holo forms. The LC-MS spectra recorded for apo and holo proteins, together with expected masses are listed in Supplementary Figure S.2. An acetyl group was used as a natural substrate mimic, which was not feasible for chemical synthesis.

Generally, the length of the bound substrate chain does influence the dynamics of ACPs[114]. For example, the study of the effect of fatty acid chain length on the dynamics of the spinach fatty ACPs showed that longer attached fatty acids result in faster hydrolysis. However, the differences in the rate of hydrolysis for the smaller fatty acids were generally small. Interestingly, however, pentadecanoic (15:0) fatty acid is the

longest fatty acid that can be easily accommodated within the spinach ACP binding pocket as longer fatty acids lead destabilization of the protein fold [74]. This suggests that using a short acyl chain as a substrate mimic should not significantly alter the dynamics and the properties of Bamb_5919 ACP compared to its natural substrate.

The transfer of the acetyl group from long holo-Bamb_5919 ACP to *holo*-Bamb_5917 PCP was monitored by detecting both acetylated and deacetylated species of carrier proteins (Figures 3.6 and 3.7). In the presence of Bamb_5919 KS^0, the acetyl group was transferred from long holo-Bamb_5919 ACP to holo-Bamb_5917 PCP, which can be seen as two peaks representing both acetylated and deacetylated long Bamb_5919 ACP and Bamb_5917 PCP, since the acetylation and deacetylation reactions are expected to be in equilibrium (Figure 3.6A). In a negative control experiment, where Bamb_5919 KS^0 was omitted from the reaction mixture, the transfer of the acetyl group was not observed and only long acetyl-holo-Bamb_5919 ACP and holo-Bamb_5917 PCP were detected (Figure 3.6 B). In addition to confirming the result obtained by Dr. Sydor, this also ensured that Bamb_5919 KS^0 and Bamb_5917 PCP are functional.

The same acetyl transfer reaction was monitored for short Bamb_5919 ACP. Both acetylated and deacetylated forms of holo-Bamb_5919 ACP and holo-Bamb_5919 PCP were observed, which confirmed that the truncated Bamb_5919 ACP retained its bioactivity (Figure 3.7).

Thus, the short Bamb_5919 ACP was used for all the following experiments and is referred to as just Bamb_5919 ACP further in the text.

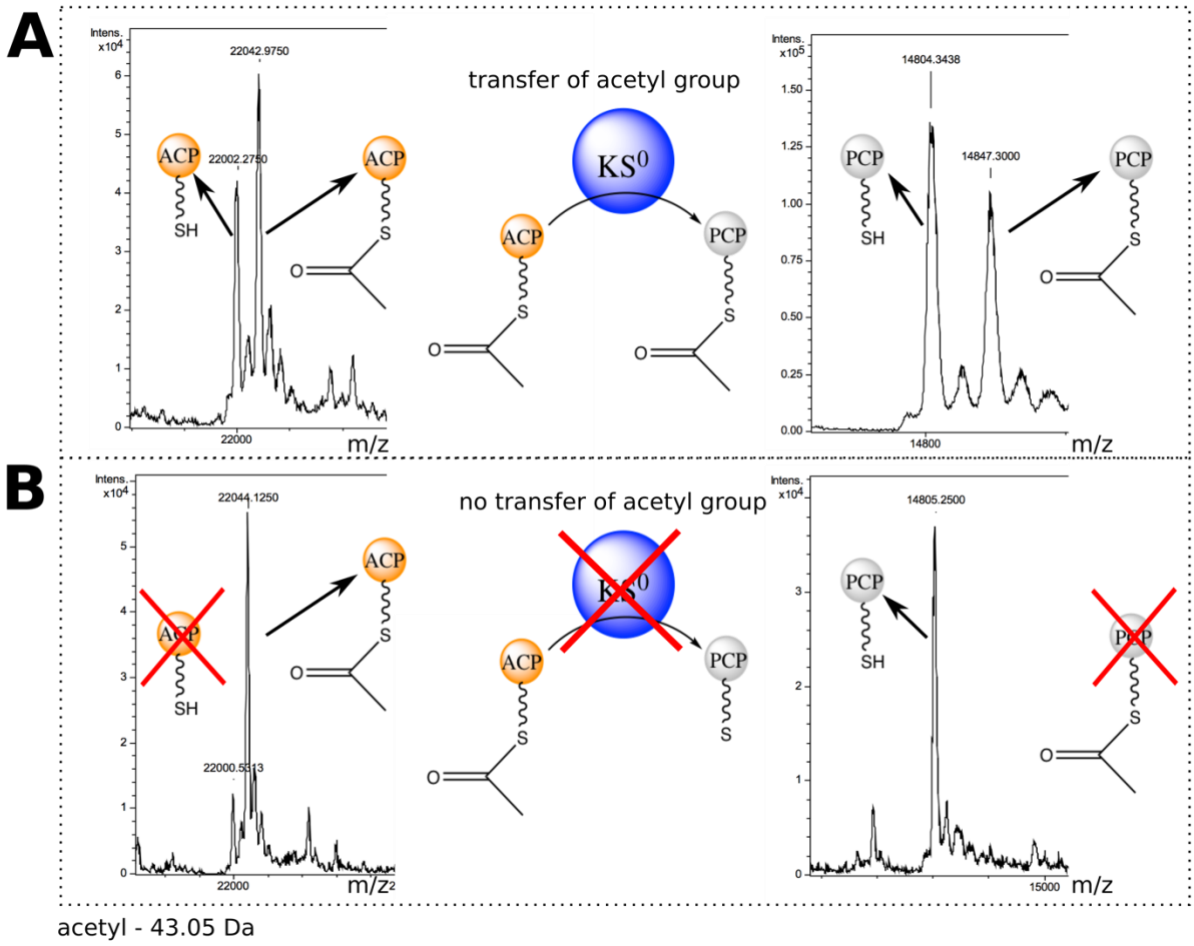


Figure 3.6 LC-MS analysis of the bioactivity assay for long Bamb_5919 ACP. The shown deconvoluted LC-MS mass spectra represent long holo-Bamb_5919 ACP and Bamb_5919 PCP in presence (A) and absence of KS^0 B). The expected masses of the used proteins are listed in Supplementary Figure S.2. Conversion from apo to holo form results in the mass gain of 358 Da, as a result of addition of phosphopantetheine. The reaction buffer was equivalent to NMR buffer.

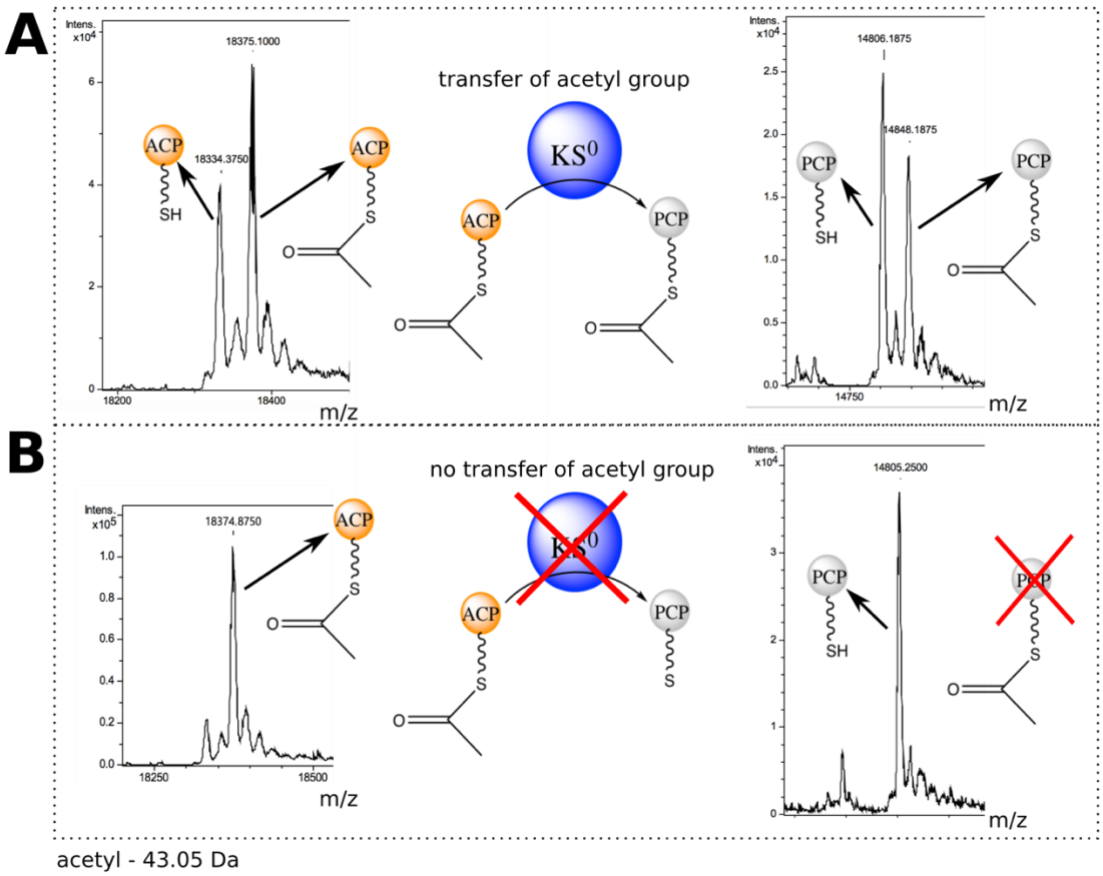


Figure 3.7 LC-MS analysis of the bioactivity assay for long Bamb_5919 ACP. The shown deconvoluted LC-MS mass spectra represent short holo-Bamb_5919 ACP and Bamb_5919 PCP in presence (A) and absence of KS^0 B).

3.2.5 Kinetic profile of Bamb_5919 ACP:Bamb_5917 KS^0 interaction

The kinetics of the interaction between Bamb_5919 ACP and Bamb_5919 KS^0 was investigated using bilayer interferometry (BLI) and isothermal calorimetry (ITC).

Like SPR and ITC, BLI is a label-free optical analytical technique for measuring biomolecular interactions. It analyses the interference pattern of white light reflected from two surfaces: a layer of immobilized protein on the biosensor tip and an internal reference layer[115–117]. Any change in the number of molecules bound to the biosensor tip causes a shift in the interference pattern that can be measured in real-time. BLI provides the ability to monitor the kinetics of a molecular reaction by monitoring rates of

association and dissociation (k_{on} and k_{off} rates, respectively) using simple dip and read assays. From these rates, the overall reaction affinity (K_d) can be calculated.

Kinetic equilibrium association (K_d) and dissociation (K_a) constants are defined as:

$$K_d = \frac{k_d}{k_a} = \frac{[A][B]}{[AB]} \qquad K_a = \frac{k_a}{k_d} = \frac{[AB]}{[A][B]},$$

where A and B are the ligand and analyte concentrations and k_a/k_d are the on- and off-rates. K_d is represented in molar units (M). The K_d is the concentration of the protein which occupies half of the binding sites of the binding partner at equilibrium.

Therefore, a lower K_d value is characteristic of stronger protein-protein interactions since less free protein is required to occupy 50% of the binding sites. Reversely, a high K_d value implies a weak binding. K_a is the reciprocal of K_d and measures the affinity of the ligand for the target. A ligand with a high K_d value has a low K_a value and therefore low affinity for the target.

In biosensor analysis, one interacting partner is immobilized on the surface of the biosensor (ligand) and the other remains in solution (analyte). The assay begins with a baseline or equilibration step using assay buffer (Figure 1.8). Next, a ligand molecule is immobilized on the surface of the biosensor (loading). Biosensors are dipped into the buffer solution for a baseline step after ligand immobilisation and to assess the ligand's load level. Next, in the association step, biosensors are dipped into a solution containing the binding partner (the analyte, where the binding of the analyte to the ligand being immobilised on the sensor surface is quantified. In the final step, the biosensor is dipped into a buffer solution which does not contain an analyte so the bound analyte can dissociate from the ligand (dissociation).

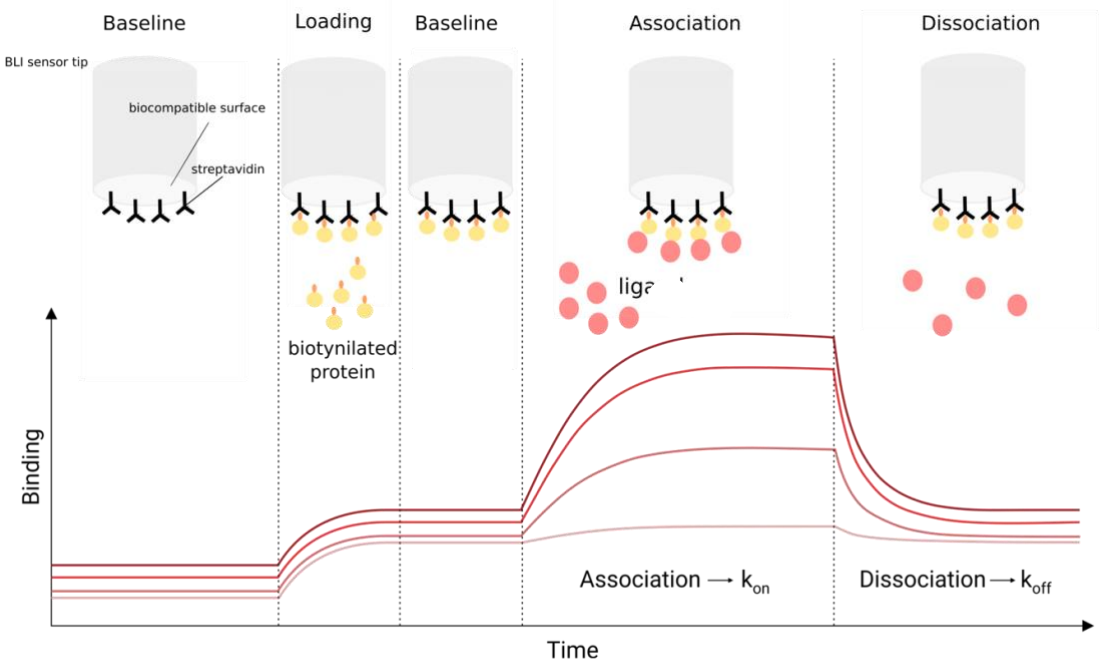


Figure 3.8 Schematic representation of a binding kinetics experiment on streptavidin biosensors. The sensorgram shows loading, baseline, association, and dissociation assay steps.

Having a precise K_d value for Bamb_5919 ACP:Bamb_5919 KS^0 binding was important to complement solution-state NMR titration analysis, which is discussed later in the chapter. The buffer conditions selected for the affinity measurements were thus the same as those selected for NMR titrations. Apo and holo-Bamb_5919 ACP were bound to the sensors and titrated with increasing concentrations of Bamb_5919 KS^0. The binding was evident from the increase in signal at higher concentrations of Bamb_5919 KS^0 (Supplementary Figure S.3). Fitting of the data to a direct binding isotherm revealed a K_d of 2.43 ± 0.48 µM (an average value of the triplicate measurements) for the binding of *apo*-Bamb_5919 ACP and 0.16 µM ± 0.02 µM for the binding of holo-Bamb_5919 ACP (Figure 3.9), indicating that *holo* form of the protein binds stronger to the ligand compared to *apo* form. This was consistent with a study of type 1 mycolactone PKS, where *holo-ACP* from module 5 binds stronger to KS partner compared to *apo*-ACP[118].

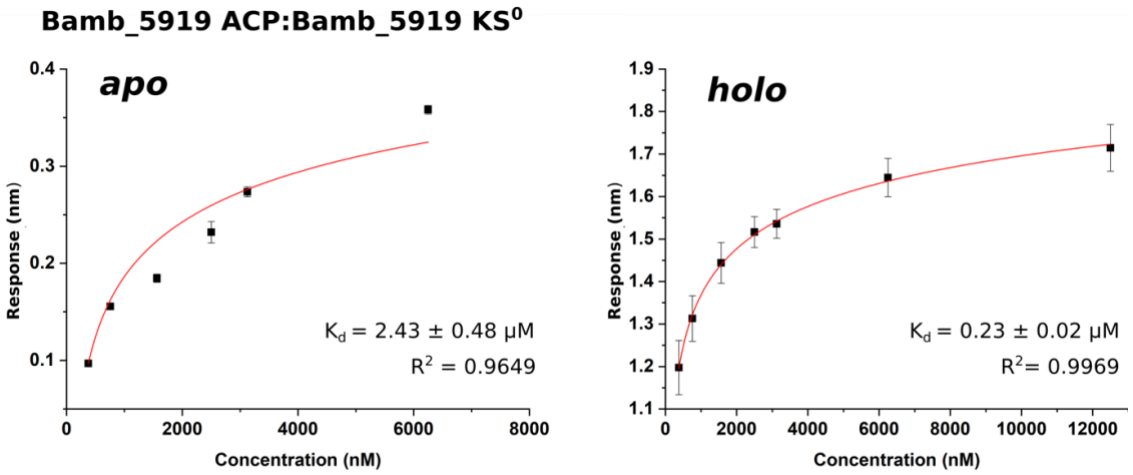


Figure 3.9 BLI steady-state analysis profile for of a Bamb_5919 ACP-Bamb_5919 KS^0 interface. The steady-state analysis determines the affinity constant K_d from the measures equilibrium response. A The binding of apo- Bamb_5919 ACP with Bamb_5919 KS^0. B) The binding of holo-Bamb_5919 ACP to Bamb_5919 KS^0. The raw binding data and fitting curves are shown in Supplementary S.3. The data presented in the Figure is an average of the triplicate measurements. Errors resulted from the differences between observed responses in each dataset.

However, the fitting of the real-time sensorgram curves showed the characteristics that deviated from the classical 1:1 bimolecular interaction. The best fit was achieved with 2:1 fitting, which indicated complex, non-ideal binding. Ideally, assay conditions should be optimized until the profile displays 1:1 binding characteristics. A complex, non-ideal binding profile in a kinetic assay can be caused by several factors, such as non-specific binding to the biosensor, when analyte concentrations are far above the K_d or when the ligand density on the surface is high. However, lowering the analyte concentration range in this assay, as well as lowering the ligand density on the surface, did not significantly shift the fitting to 1:1 binding scenario. Thus, the non-ideal binding was probably the result of the structural heterogeneity of the ligand or analyte, occurring either in the original sample or as a result of artefacts in the assay, or the presence of partially active or inactive proteins.

BLI provided K_d values of the interactions, but to get information about changes in enthalpy of the binding, ITC was performed. ITC works by directly measuring the heat that is either released or absorbed during a biomolecular binding event. ITC is the only

technique that can simultaneously determine K_d, enthalpy and entropy of binding in a single experiment.

Enthalpy and entropy of binding, together with Gibbs energy, form the basic thermodynamic binding profile for a specific interaction. Gibbs energy is calculated from K_a:

$$\Delta G = -RTlnK_a = RTlnK_d,$$

Where T stands for absolute/thermodynamic temperature and R for gas constant. Furthermore, Gibbs energy can be partitioned into its enthalpic, ΔH, and entropic, TΔS, contributions:

$$\Delta G = \Delta H - T\Delta S.$$

Combination of these three thermodynamic parameters offers basic and valuable information on the binding process. For example, endothermic profile suggests that protein-protein interactions are mediated by hydrophobic residues whilst exothermic reaction indicates polar interactions[114].

As well as for BLI, no tagging or modifications of binding partners are required. However, contrary to BLI, it does not require protein immobilization so ITC measures the affinity of binding in the native states of binding partners. However, the disadvantages of ITC include a large amount of required sample and the hardly detectable small binding enthalpies exhibited by non-covalent complexes, such as the complex formed by carrier proteins and ketosynthases.

During ITC measurement, a biomolecule solution is initially placed in the sample cell during the measurement and a ligand solution is placed in the syringe in a corresponding buffer. When the ligand solution is injected into the cell and the interaction occurs, the ITC instrument detects the released or absorbed by measuring the changes in the power required to maintain the same temperature between the reference and the sample cell (DP). Injections are performed repeatedly and result in peaks that become smaller as the biomolecule becomes saturated. Eventually, the peak sizes remain constant and represent only the heats of dilution[114].

The binding of apo and holo-Bamb-5919 ACP and Bamb_5919 KS^0 was evident from the decrease in DP (Figure 1.10). A single measurement of the interaction recorded for apo-Bamb_5919 ACP indicated that the binding is weaker compared to holo-Bamb_5919 ACP, which was consistent with the BLI results. However, the attempts to replicate the measurements failed due to the presence of bubbles in the samples that affected the measurements, even after degassing. This indicated that the sample solutions were highly viscous, which complicated the removal of the bubbles. Also, the offset (signal obtained from the reference analyte buffer to water titrations) was substantial and some of the replicated measurements could not yield distinguishable signal originating from the very binding reaction. This background signal probably emerged from the buffer impurities, which can affect the measurements since the signal originating from the mixing of these impurities with water can generate heat which can be greater than the heat released or absorbed by the protein interaction. Furthermore, the stoichiometry obtained from the measurements was far off 1:1 binding scenario. This indicated that actual concentrations of the proteins were not correctly estimated, probably due to adverse effects of an increase in temperature during the transport of Bamb_5919 KS^0.

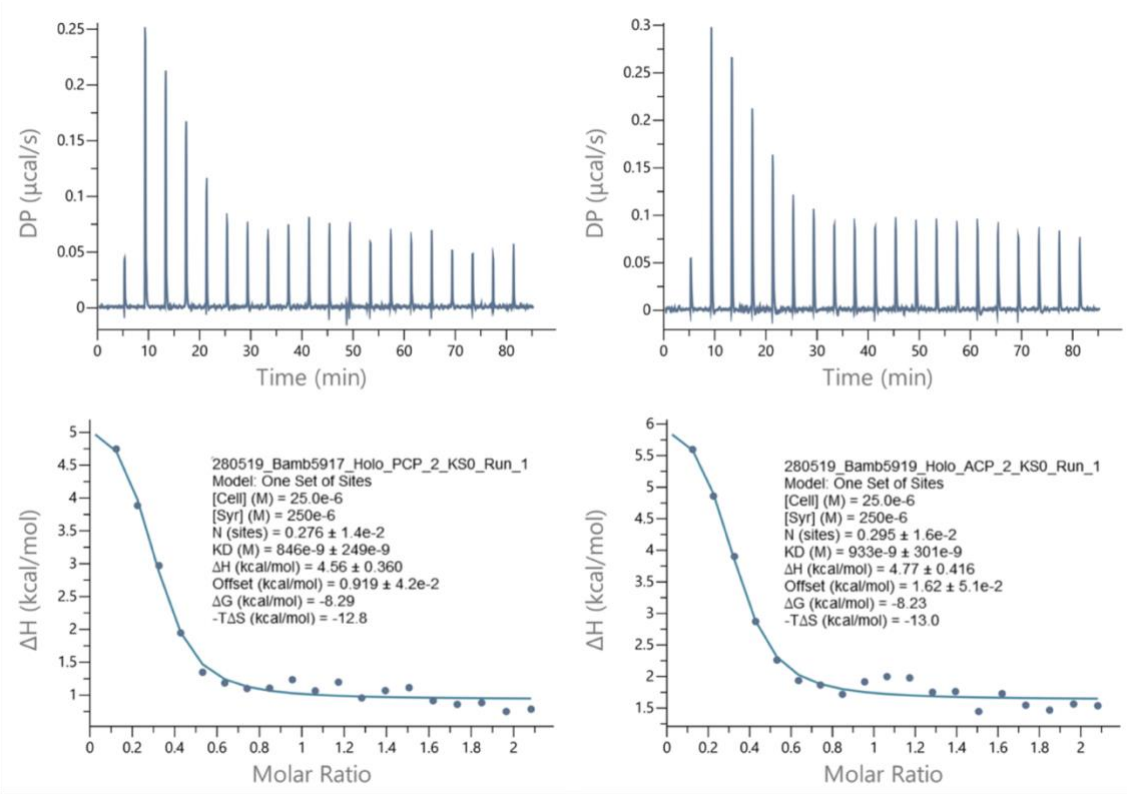


Figure 3.10 ITC analysis of the binding between Bamb_5919 ACP and Bamb_KS^0. The measurement obtained for apo-Bamb_5919 ACP (upper and lower left figure). B) The results obtained for two replicates of apo-Bamb_5919 ACP measurements (upper and lower right figure). The dilution control has been subtracted from titration.

3.3 NMR spectroscopy for the study of proteins

To understand the NMR experiments used to analyse the protein interactions, the important concepts related to NMR spectroscopy, will be described first[114].

NMR is a powerful tool for studying the structures of macromolecules and more than 12,500 NMR structures of proteins have been deposited in the PDB to date. The development of Fourier transform NMR and multidimensional NMR methods allowed for faster experiments and correlation of resonances between spatially proximal or bound nuclei, which improved the applicability of NMR to study proteins. NMR can be used to produce structures of proteins up to ~30 kDa using the standard methods that will be described in this chapter. Advances in molecular biology methods have enabled isotopic enrichment of ^{13}C and ^{15}N nuclei that are relevant for NMR. Overexpression of proteins in *E. coli* has made sequential backbone assignment of backbone $^{13}C\alpha$, $^{13}C_{CO}$, ^{1}H and ^{15}N atoms possible. NMR pulse sequences, such as TROSY[114,119], together with developments

in data processing have allowed for even larger proteins to be studied (50-1000 kDa). Proteins of that size are otherwise not accessible to high-resolution solution-state NMR[120]. The assignment of the protein backbone resonances can also help in the resonance assignment of the side-chain atoms directly bound to the backbone residues and thus enable the characterisation of the dynamics and conformational states of both the protein backbone and sidechain[121,122]. Furthermore, the development of high field NMR spectrometers with ^{1}H Larmor frequencies of 900 MHz and above, together with cryogenically cooled probes, have significantly increased the sensitivity and resolution of NMR experiments. NMR can be used to study proteins in solution and so it offers a significant advantage over alternative techniques in maintaining natural protein dynamics and flexibility. NMR can also be used to map protein-protein interactions and report on the timescales of internal dynamics, conformational exchange and ligand binding.

The NMR signal

All nuclei which do not contain an even number of protons and neutrons possess the quantum mechanical property known as the spin angular momentum, commonly referred to as the 'nuclear spin', which is described using the nuclear spin quantum number, I. Spin is an intrinsic property of the nucleus, like mass and charge, and does not arise from its spinning. The nuclei with I = ½, such as ^{1}H, ^{15}N and ^{13}C are the most useful for protein NMR because they split into only two energy levels that are parallel and antiparallel to the applied magnetic field. The ratio of parallel to antiparallel spins is given by a Boltzmann distribution which shows that the two energy levels are nearly equally populated at thermal equilibrium. As a result, NMR possesses low sensitivity. ^{1}H is naturally abundant (99.99 %), whilst ^{13}C (~1 % natural abundance) and ^{15}N (~0.4 %) can be incorporated during protein expression to increase the abundance. The nuclear spin angular momentum of a nucleus with $I \neq 0$ gives rise to a magnetic moment. The z-component of the magnetic moment (μ) is the z-component of the angular momentum ($\hat{I}_z$) multiplied by the gyromagnetic ratio (γ), which is a fundamental property of the nucleus:

$$\mu = \gamma \hat{I}z.$$

The energy of the interaction depends upon the angle (θ) between the magnetic moment and the applied field. The energy is lowest when the applied field and magnetic moment are parallel (θ = 0) and highest when they are antiparallel (θ = π radians). Therefore, the alignment of the nuclear magnetic moment to the applied field is energetically favourable and the lower energy orientations become preferentially populated over time. The energetic preference is opposed by thermal motion so the net alignment that arises is very small. Nonetheless, even this small alignment is such that the sample acquires its magnetic moment, which is known as the bulk or equilibrium magnetisation. This alignment is in the same direction as the applied magnetic field, i.e. along the z-axis. The extent of the bulk magnetisation increases with the increase in the field strength. For protein NMR, magnetic fields of more than 500 MHz are required to achieve sufficient magnetisation. The equilibrium magnetisation does not change its direction or size over time. However, if this magnetic moment is perturbed away from the z-axis towards the *xy*-plane, the vector precesses about the z-axis (Figure 3.11). The frequency of precession depends upon the gyromagnetic ratio (γ) and the applied field (B_0) and is known as the Larmor frequency (ω_0):

$$\omega_0 = -\gamma B_0.$$

The NMR signal originates from the precession of the magnetisation vector, which is detected by a coil of wire placed around the sample along the x-axis (Figure 3.11). When the magnetisation vector crosses the coil, the current is induced and recorded as the free induction decay (FID). The signal decays due to relaxation of the magnetisation vector back to equilibrium.

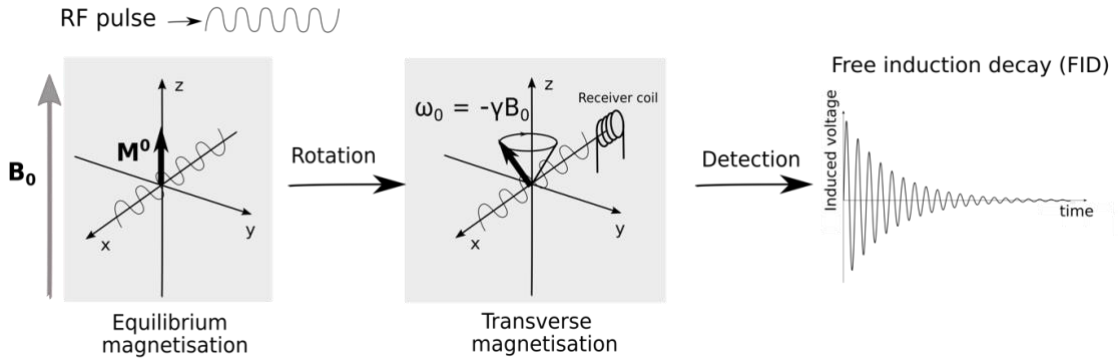


Figure 3.11 Magnetisation and the NMR Signal. The equilibrium magnetisation that builds up after a few seconds in a magnetic field is along the z-axis. The direction and magnitude of the equilibrium magnetisation vector do not change over time. Following perturbation of the magnetisation vector away from the z-axis into the xy-plane, the vector precesses about the z-axis at the Larmor frequency (ω0 = -γB0) for a nucleus with a positive gyromagnetic ratio such as a proton. As the vector crosses the receiver coil placed around the x-axis, an electrical current is generated. The current is known as the free induction decay (FID) and represents an NMR signal.

The precise magnetic field experienced by a nucleus depends on its chemical environment. Each nucleus within a protein effectively experiences a different B_0. Therefore, the Larmor frequency of each nucleus is also different. This can be described mathematically by incorporating an additional term into the equation for the Larmor precession known as the chemical shift:

$$\omega_0 = -\gamma(B_0 - B_i),$$

where B_0 is the external magnetic field and B_i - induced magnetic field.

The FID, which is converted to a spectrum of frequencies using the Fourier transform, contains the precession frequency of each spin active nucleus in the protein. These frequencies are given relative to the receiver reference frequency, which is typically placed by the user in the centre of the frequencies of interest. For example, to detect the peaks between 0 and 1000 Hz, the receiver would be placed at 500 Hz resulting in a spectrum going between -500 and +500 Hz. Further to this, the frequencies (v) are normalised relative to the frequency of a reference compound and quoted as chemical shifts (δ) in parts per millions (ppm), to remove their dependence on the applied field:

$$\delta\ (ppm) = \ 10^{-6} \cdot \frac{\nu - \nu_{ref}}{\nu_{ref}}.$$

The final NMR spectrum contains a peak for each spin active nucleus in the protein at its Larmor frequency, given as a chemical shift in ppm.

The NMR experiment

The rotation of the equilibrium magnetization away from the z-axis and into the xy plane is fundamental to the NMR signal. Switching off the applied field (B_0) along z and replacing it with a field along x or y is the most obvious means of achieving this, but it is simply not possible to switch the powerful superconducting magnets used to establish the bulk magnetisation off and on in this way. However, the applied field can be overcome by an additional small, oscillating magnetic field (B_1) applied along the x-axis (or y-axis), known as the radio frequency (RF) field (or pulse). The B_1 field is generated using the detector coil and is set to linearly oscillate at or close to the Larmor frequency. It can be considered as two counter-rotating fields (B_{1+} and B_1), rotating at a frequency of ω_1. The rotating field significantly interacts with the Larmor precession of a nucleus if it is rotating in the same direction but has no significant interaction if rotating in the opposite direction. This means that one of the two components can be ignored. For example, for a nucleus with a negative gyromagnetic ratio such as ^{15}N, only the B_{1+} component of the RF field is rotating with the same sense as the Larmor frequency and so the B_{1-} component can be ignored. Now, by observing B_{1+} from a rotating frame that rotates with the same sense and frequency as B_{1+}, B_{1+} appears static.

Observing the Larmor precession in this same rotating frame causes an apparent change to the Larmor frequency. The apparent Larmor frequency in the rotating frame is known as the offset (Ω) and is given by:

$$\Omega = \ \omega_0 - \ \omega.$$

The Larmor frequency is related to the applied magnetic field and so, in the same way, the offset is related to the apparent field experienced in the rotating frame (ΔB):

$$\Omega = -\gamma \Delta B.$$

This proportional relationship between the apparent field and the offset allows the apparent field to be reduced to zero simply by setting the frequency of the rotating frame ($\omega_{rotating\ frame}$) equal to the Larmor frequency (ω_0). The rotating frame, therefore, gives us the power to apply an oscillating magnetic field along x or y that appears stationary as well as effectively converting B_0 into a reduced field (ΔB) that is very small. The field now experienced by each spin is a vector sum of the apparent field (ΔB) and the RF field (B_1) and is known as the effective field (B_{eff}). When the frequency of the RF field is selected to match the Larmor frequency exactly, Ω = 0, ΔB = 0 and so $B_{eff} = B_1$. This is called an 'on-resonance' pulse and it causes the magnetisation to rotate about the axis to which it was applied. The angle through which the magnetisation vector rotates (flip angle, β) depends on the frequency of the RF pulse and pulse length (t_{app}):

$$\beta = \omega 1 t_{app}.$$

When the frequency of the RF field is very different from the Larmor frequency, $\Omega \approx \omega_0$ and so B_{eff} is dominated by ΔB. This is called an 'off-resonance' pulse and it essentially does not affect the magnetisation. The Larmor precession of ^{1}H, ^{15}N and ^{13}C differ significantly due to their different gyromagnetic ratios and so a pulse that is on-resonance with protons will not affect the other nuclei. The basic 'pulse-acquire' NMR experiment consists of a 90° pulse on the x-axis, on resonance with the nucleus of interest, followed by signal acquisition (Figure 3.9). In terms of the 'vector model', the 90° pulse on x simply rotates the magnetisation vector from z onto -y and the magnetisation then precesses in the xy-plane during the acquisition of the FID. The addition of a 180° pulse before acquisition turns the pulse-acquire into the 'spin-echo'. In the spin-echo, the equilibrium magnetization is rotated into the transverse plane as in the pulse-acquire and the spins precess within the xy-plane for time, τ. The spins do not precess at identical frequencies and so after τ, the magnetisation of each spin will have reached a different point along a circular path around z. When a 180° pulse is applied, all of the spins are rotated to be exactly opposite where they were before the pulse. The spins continue to precess at the

same frequency and in the same direction as before and so after another period of exactly τ, they will be aligned on the y axis. The chemical shift is said to be 'refocused'.

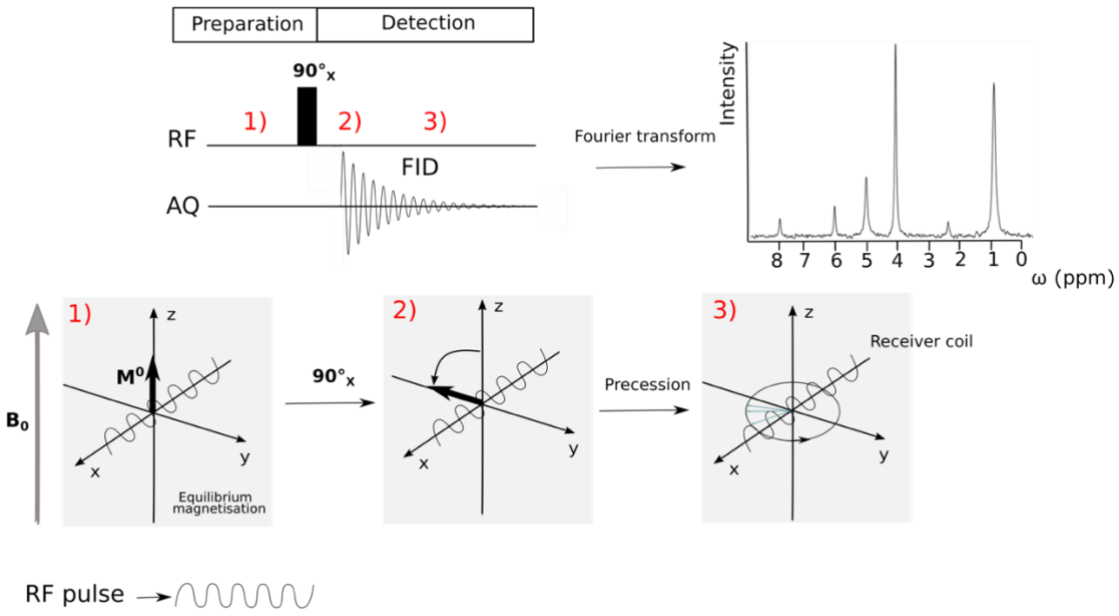


Figure 3.12 The spin echo and pulse-acquire. The RF line contains the information on the RF pulses and the acquisition (AQ) line represents when the signal is recorded. The acquired FID is processed using the Fourier Transform to produce a spectrum with a peak for each spin active nucleus at its Larmor frequency, expressed as the chemical shift in ppm. The vector model diagram illustrates the changes to the bulk magnetisation that occur at points (1), (2) and (3) during the experiment shown.

Multi-dimensional NMR

A good quality heteronuclear single quantum coherence experiment (HSQC) is a prerequisite for successful protein assignment. The spectrum comprises a set of peaks representing every NH group in the protein (Figure 3.13).

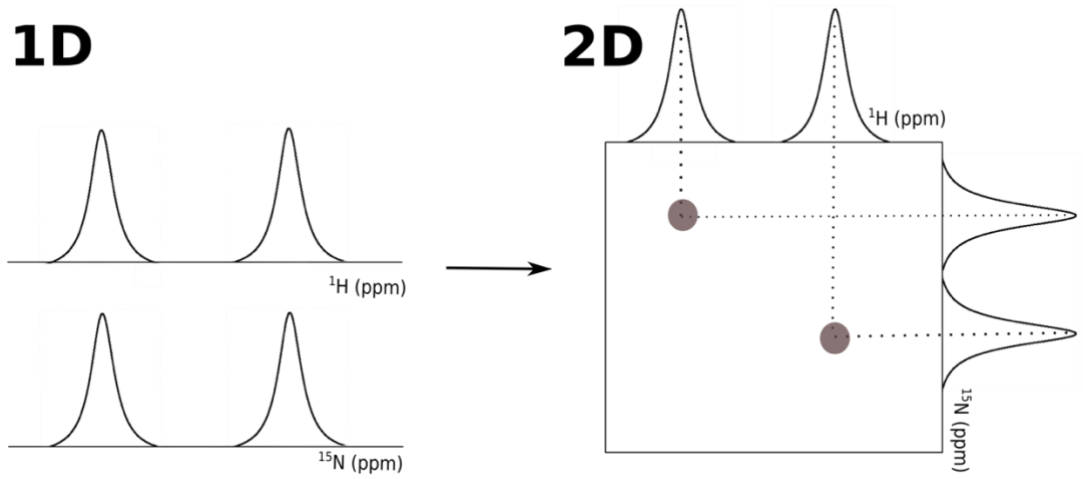


Figure 3.13 Simplified and schematic representation of a 2D experiment. 2D experiments can be created by combining two 1D ^{1}H and ^{15}N spectra. Peaks in the 2D spectrum emerge only if there is a correlation between peaks in the 1D spectra. This can happen through a direct bond or through space.

This includes the backbone NH group of every residue (absent in prolines), the sidechain NH of tryptophan, histidine and arginine, the sidechain NH_2 of arginine, glutamine and asparagine and the sidechain NH^{3+} of lysine. For arginine, however, the NH and NH_2 groups are often very weak due to exchange broadening and so may not be observed. The NH of histidine and the NH^{3+} groups of lysine are also not observed. The NH_2 groups of glutamine and asparagine, however, give stronger peaks in the top right-hand region of the HSQC (δ_N = 110-115 ppm and δ_H =6.5-8.0). Two peaks are observed with equivalent nitrogen shifts, representing each proton attached to single nitrogen. The N-terminus is usually not visible due to exchange broadening. The ^{1}H and ^{15}N chemical shifts of each peak are determined by the chemical environment of the amide group, therefore the HSQC spectrum depends upon the fold of the protein. For this reason, the HSQC spectrum is often referred to as the protein's fingerprint and a well-dispersed spectrum

is characteristic of a well-folded protein. The version of an HSQC experiment that was used in the work presented in the thesis is called HMQC.

Also, 2D NOESY and TOCSY can be used to report on the secondary structure of a protein. The resolution of the peaks in the HN-Hα region of the TOCSY can indicate whether the protein can be studied by NMR. If the peaks in this region are very broad it could indicate oligomerisation or aggregation of the protein,that the protein is too large for classical NMR studies and that further NMR experiments will not provide the necessary resolution for structural studies. Good resolution in this region indicates that the much higher resolution double and triple resonance ($^{1}H/^{15}N$ or $^{1}H/^{15}N/^{13}C$) experiments will be of ample resolution for backbone and sidechain assignments and structural studies.

The INEPT[123] is an important building block for many NMR experiments, including the HSQC. In its simplest form, this contains two INEPT sequences separated by an evolution time, t_1 (Figure 1.10). During t_1, the proton chemical shift is refocused by a 180° pulse but the ^{15}N chemical shift is allowed to evolve.

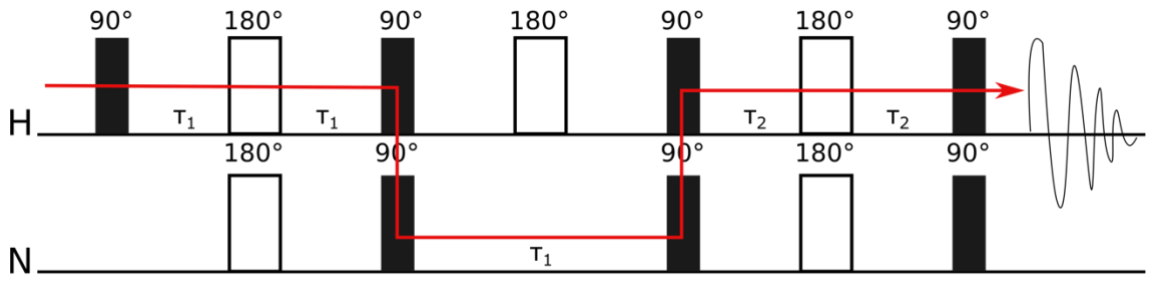


Figure 3.14 INEPT pulse sequence. The evolution of the magnetisation during pulses and periods of free precession. The red arrow indicates the transfer of magnetisation from the proton to ^{15}N. This is necessary to encode the ^{15}N shifts in the final signal. Thus, the ^{15}N chemical shifts are encoded and the magnetisation has been transferred back to the proton.

The final signal encodes the ^{15}N chemical shifts modulated by t_1 and is detected on the proton channel. The resulting ^{15}N-HSQC spectrum contains a signal for every NH group in the protein. Further extensions to the HSQC include the triple resonance experiments used for protein backbone assignments, in which the magnetisation is selectively transferred using scalar couplings between protons, ^{15}N and ^{13}C nuclei.

Multi-dimensional NMR experiments that correlate the resonance frequency of the backbone nuclei of one amino acid residue (i residue) with that of the preceding (i_{-1}) or the succeeding (i_{+1}) residue are typically used for the sequential assignment of the protein backbone and sidechain nuclei ($^{12}C\alpha$, $^{12}C\beta$, ^{13}CO, ^{1}H, and ^{15}N).

For example, the 3-dimensional HNCA experiment provides a correlation between the ^{1}H and ^{15}N nuclei of the backbone amide group with the $^{13}C\alpha$ chemical shifts of the i residue and the i_{-1} residue. The connectivity information can be used in a chain process (linking the i_{-1} residue with the i residue and the i residue with the succeeding i_{+1} residue etc.) to complete assignment of the backbone ^{15}N, ^{1}H, and $^{13}C_{\alpha}$ nuclei. To distinguish possible ambiguities on which resonances originate from the i and i_{-1} residue, HN(CO)CA experiment can be performed, which only shows the correlation of the backbone ^{1}H and ^{15}N with the i_{-1} $^{13}C_{\alpha}$ nuclei and not the $^{13}C_{\alpha}$ resonance. Other experiments, such as the CBCA(CO)NH and the HNCACB experiments (which provide similar correlation information as the HN(CO)CA and the HNCA experiments with the addition of $^{13}C\beta$ resonances) or the HNCO experiment (which provide correlation of the backbone ^{1}H and ^{15}N resonances with the ^{13}CO resonance) can aid in the assignment of the protein backbone nuclei. The assignment of additional sidechain nuclei can be performed using the hCCH-TOCSY experiment that shows the correlation of resonances of the sidechain ^{13}C atoms to the backbone nuclei ($^{13}C_{\alpha}$ and $^{1}H_{\alpha}$).

The NMR resonance assignments for backbone and sidechain atoms can be used for the study of both structural as well as dynamics of the protein. Chemical shifts reflect the overall chemical environment of the nuclei and changes in the backbone ^{1}H, ^{13}C, and ^{15}N chemical shifts can reflect changes in the protein conformation, presence of bound ligand, or differences in side-chain due to mutation.

3.3.1 2D NMR analysis of *apo*-Bamb_5919_ACP

The ^{1}H-^{15}N HMQC spectrum recorded for *apo*-Bamb_5919_ACP was well dispersed (Figure 3.15A.), indicating a well-folded protein. The ^{1}H–^{15}N HSQC of the Bamb_5919 ACP domain was therefore expected to contain a total of 142 backbone NH peaks. The initial number of the picked peaks was smaller than expected. However, without 3D experiments, it was not possible to distinguish overlapped peaks. Also, 2D ^{1}H–^{1}H NOESY (with ^{15}N decoupling) was recorded on *apo*-Bamb_5919 ACP. In NOESY, there were many HN-HN cross-peaks in the amide region, arising from through-space NOE transfer of magnetisation between adjacent NH groups (Figure 3.15B). The abundance of cross-peaks in this region indicated the expected well-folded structure. Together, these spectra indicated that Bamb_5919 ACP domain is most likely assignable. Hence, ^{15}N and ^{13}C labelled protein was expressed and purified for use in the triple resonance experiments required for backbone and sidechain assignment. The LC-MS spectra of the labelled constructs are presented in the Supplementary Figure S2.

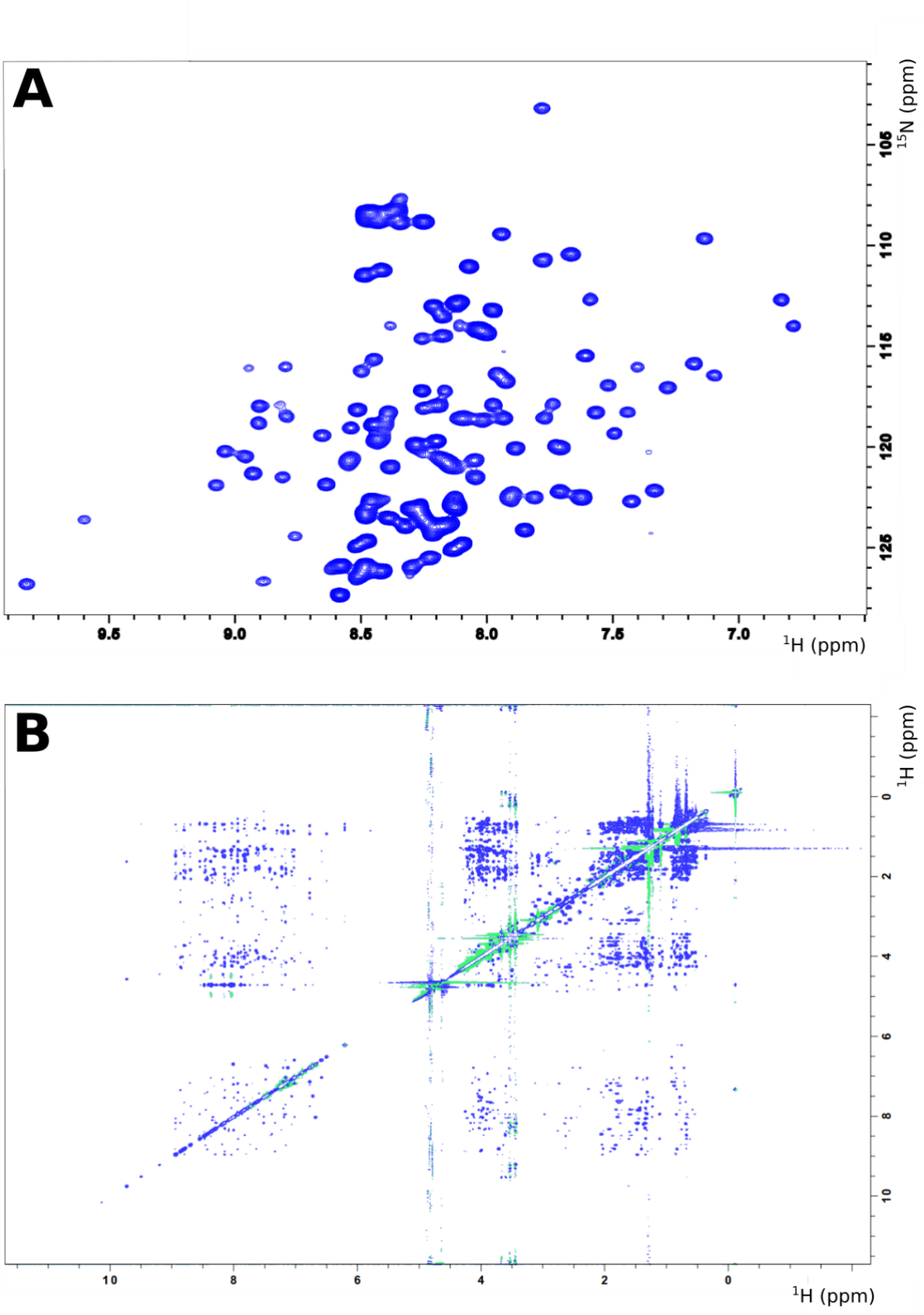


Figure 3.15 2D NMR analysis of apo-Bamb_5919 ACP. ^{1}H-^{15}N HMQC (A) and ^{1}H-^{1}H NOESY (B) recorded at 700 MHz, 298 K.

3.3.2 Backbone assignment of apo-Bamb_5919 ACP

To study conformational and dynamic changes in the Bamb_5919 ACP of upon interaction with Bamb_5919 KS^0, an assignment of the protein backbone residues was required. Due to the size of the Bamb_5919 ACP (~ 15 kDa), 3D heteronuclear triple resonance experiments, were used to sequentially assign the protein backbone. Such an assignment relies upon a series of through bond experiments to be analysed with reference to the HSQCs. The amide chemical shift of each residue is encoded in the ^{15}N-HSQC. ^{15}N-HSQC is the reference point during the assignment process. The navigation between the 3D spectra is done by using strip plots. (Figure 3.16).

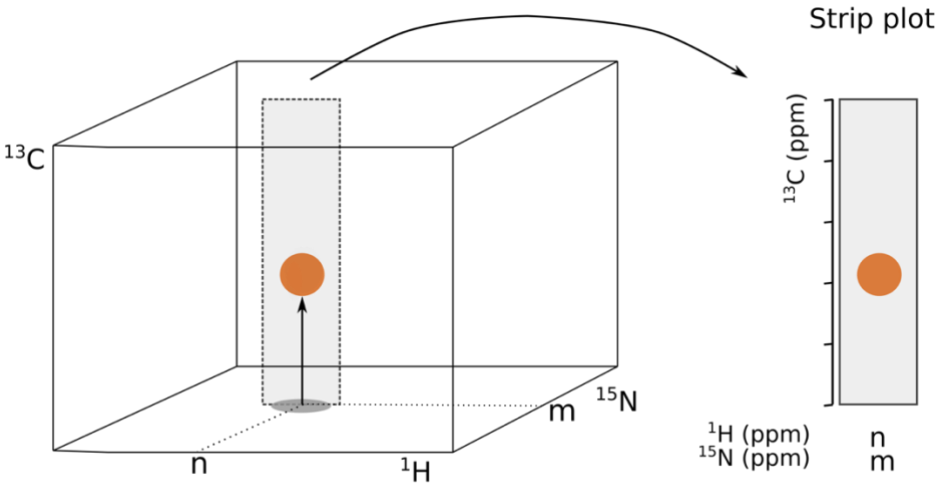


Figure 3.16 Representation of a 3D experiment. An HSQC peak is extended into a 3D carbon dimension to form a 3D dataset. Such a dataset can be analysed by using strip plots. Strip plots represent a ^{13}C dimension at a ^{1}H-^{15}N position. This way, it is possible to distinguish peaks that are overlapped in HSQC, which is often the case in proteins.

This section will give a brief introduction to the NMR experiments used for the amino acid assignment and the sequential assignment method.

The backbone assignment was performed using serial triple resonance experiments: HNCO, HNCA, HNCACB and HN(CO)CACB. These experiments depend upon the transfer of magnetisation via scalar couplings between the amide proton, amide nitrogen and carbon atoms. The magnetisation transfer pathways for each experiment are illustrated in Figure 3.17.

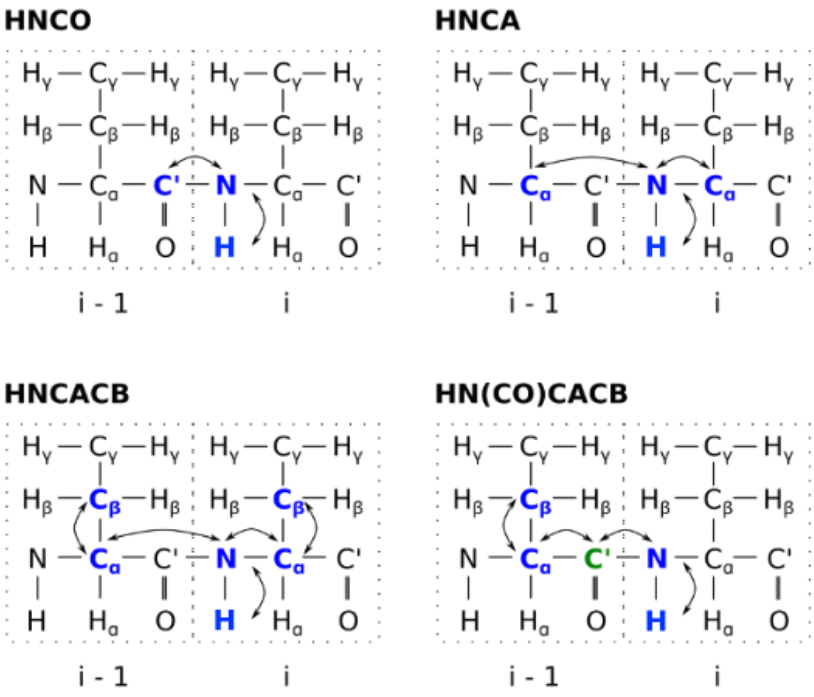


Figure 3.17 Magnetisation transfer pathways in triple resonance NMR experiments. The diagrams illustrate the transfer of magnetisation via one or two bond J couplings, indicated by double-headed arrows. In each case, the magnetisation begins and ends on the amide proton. The chemical shift is allowed to evolve on all of the nuclei highlighted in blue and, as such, the chemical shift of each is encoded. The nucleus highlighted in green are utilised for magnetisation transfer in order to restrict the transfer pathway to the i_{-1} residue but their shifts are refocused and so not encoded.

The HNCA contains a cross-peak resulting from the one-bond scalar coupling ($1JNC_\alpha$) between the NH and C_α and as such encodes the chemical shift of the NH and C_α for each residue. The HNCACB contains two cross-peaks for each NH group, one to the C_α and one to the C_β. These experiments are therefore used to assign the C_α and C_β corresponding to

each NH group. In some cases, the HNCA and HNCACB contain additional weaker peaks due to correlation of the NH to the C_α or C_β of the preceding residue via the $2JNC_\alpha$ coupling. First, the NH strip for a residue (i) can be located in the HNCA by navigation from the HSQC. The C_α of the preceding residue (i_{-1}) is visible in the HN(CO)CACB so can be used to locate the C_α in the HNCA of i_{-1} by matching the chemical shift. The corresponding peak is then located in the HSQC and linked sequentially to the starting HSQC peak. The i_{+1} strip can also be located using the Cα in the HNCA of residue i to find the corresponding C_α peak in the HN(CO)CA of residue i_{+1}. In this way, it is possible to step along the protein backbone.

Sequence-specific assignments were made with the sequential linkages by comparison of the Cα and Cβ chemical shifts with reference values. More than 7 million chemical shift values from NMR assignments are contained within an online database, the BioMagResBank[118]. These values have been used to calculate the minimum, maximum and average value for the chemical shift of each ^{1}H, ^{15}N and ^{13}C atom in specific amino acid types for *apo*-Bamb_5919 ACP and this information has been tabulated (Supplementary). The C_α and C_β chemical shifts for each residue were compared with the reference values for different residue types and so the possible identities of each residue were deduced. The different amino acid types have overlapping chemical shift profiles so a single C_α/C_β pair could not be assigned in isolation but, by matching the possible identities for stretches of sequentially linked residues with the amino acid sequence of *apo*-Bamb_5919 ACP, some sequence-specific assignments could be made. To overcome the limitations caused by C_α and C_β chemical shift degeneracy between amino acid types, two further experiments were used in addition to the triple resonance experiments: the ^{15}N-NOESY-HSQC and HCCH-TOCSY. The NOESY relies on through space NOE transfer, whilst the TOCSY transfer relies on through bond scalar couplings. In both cases, the protons are first excited and their chemical shifts recorded. For the NOESY, the magnetisation is then put back on the z axis before mixing. For the TOCSY, the magnetisation is transverse during the mixing step. It is during this mixing step that the magnetisation is transferred across all of the protons via either through space NOE transfer or through bond J couplings.

Both experiments end with a ^{15}N-HSQC and so the magnetisation is transferred back to amide protons via nitrogen before detection. As a result, only correlations between the

NH groups and other protons are recorded. The magnetisation transfer pathways for these two experiments are shown in Figure 3.18.

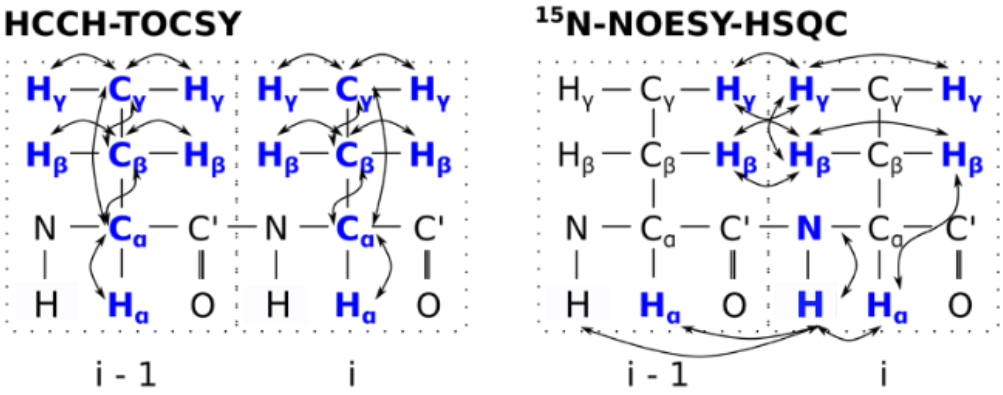


Figure 3.18 Magnetisation transfer pathways in the ^{15}N-NOESY-HSQC and HCCH-TOCSY. The nuclei for which the chemical shift is recorded are highlighted in blue.

The NOESY mixing time is selected so that only direct side-chain proton to NH group NOEs are recorded. This is typically 100 msec for small proteins. If the mixing time is too short, only the NOEs between the closest protons will be detected and longer distance information will be lost. If it is too long, the NOEs will contain additional magnetisation transferred from other nearby protons and so will not provide an inaccurate measure of the distance between sidechain protons and the amide groups. This is called spin diffusion. The TOCSY mixing time must be long enough for the magnetisation to reach the more distant protons but if it is too long the signals will be lost due to relaxation. In the TOCSY, all of the cross-peaks are intra-residue as the magnetisation is transferred through scalar couplings and is not transferred via the N-CO or C^{α}-CO bonds and the protons that are connected via the fewest bonds to the NH group will give the strongest cross-peaks. In theory, the TOCSY provides the proton shifts across the complete sidechain, providing a useful fingerprint for amino acid type. However, for most residues of *apo*-Bamb_5919 ACP, the TOCSY transfer was only sufficient to identify the H_{α}. This is because the 3JHN-H_{α} coupling constant is small (4-6 Hz) in helical proteins and so magnetisation transfer from all but the Hα requires impracticably long mixing times that would result in the pronounced signal loss to relaxation.

Conversely, the NOESY experiment relies upon through space correlations that are efficient up to 5 Å and so for *apo*-Bamb_5919 ACP it did provide sidechain fingerprints. The NOESY contains intra- and inter-residue NOEs and the intra-residue NOEs are often strong as the intra-residue protons are close in space to the NH group. The inter-residue NOEs observed in the NOESY, which may be short or long-range in terms of primary structure, can later be used to derive distance restraints for structure calculations. These increase spectral crowding but can also aid the assignment by the provision of helpful short and medium-range NOEs. For *apo* Bamb_5919 ACP, short-range NH-NH NOEs (i,i_{+1}) along the sequence and medium-range NOEs (i_{+3}) were useful in aiding the assignments. Overall, the assignment of apo-Bamb_5919 ACP relied upon HNCA, HN(CO)CACB and HNCACB. HNCO was an additional help in distinguishing peaks in the recorded ^{15}N-HMQC, especially in regions with significant peak overlap. ^{15}N-HSQC-NOESY was primarily used to assess the whether the sequential backbone assignments were correct. HCCH-TOCSY aided the assignment of residues with ambiguous or very weak Cβ shifts, especially in the folded parts of the protein. Hence, the chemical shift table presented in Supplementary contains backbone shifts obtained from HNCA, HN(CO)CACB and HNCACB only.

The pH conditions under which the NMR experiments were performed were set up to resemble the expected in vivo conditions, i.e. pH = 7. This pH is close to the pI value of Bamb_5919 ACP (pI = 7.10). However, there was no visible precipitation detected. CD analysis of Bamb_5919 ACP (Figure 3.3) showed that the protein is not negatively affected by the temperatures between 30°C and 40°C, indicating that it should not easily degrade in an NMR tube during the acquisition of 3D NMR experiments. 108 out of 142 residues were assigned (Figure 3.19). The N-terminal part of helix 2 was not assignable due to weak Cβ signals in HNCACB. Furthermore, serine residue that anchors the phosphopantetheine arm was most likely invisible in ^{15}N HMQC spectrum.
Examples of sequential walks through 3D experiments used in the assignment are shown in Figure 3.17.

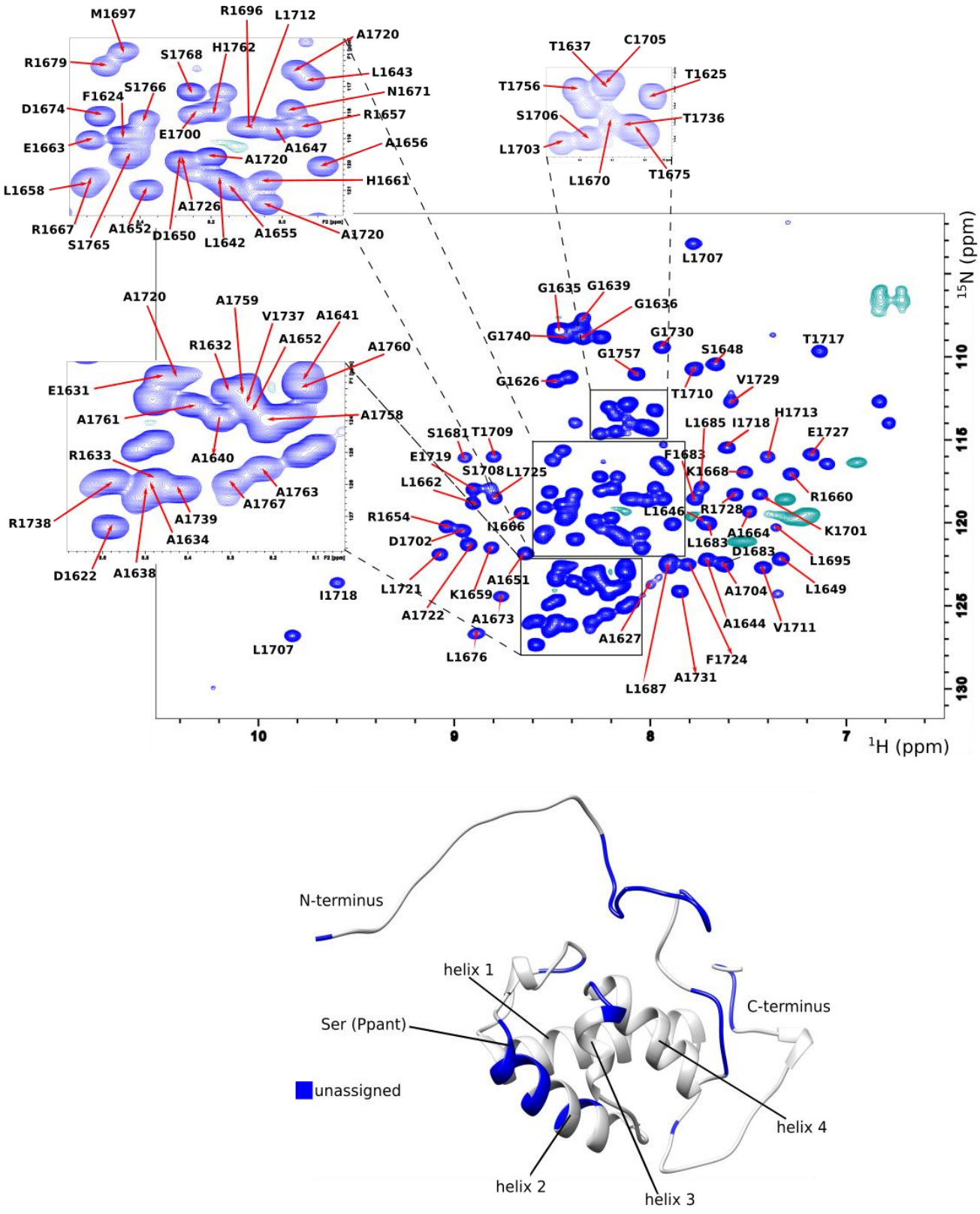


Figure 3.19 . The assigned ^{1}H-^{15}N-HSQC of apo-Bamb_5919 ACP. The residues that were unassigned are also shown mapped onto homology model.

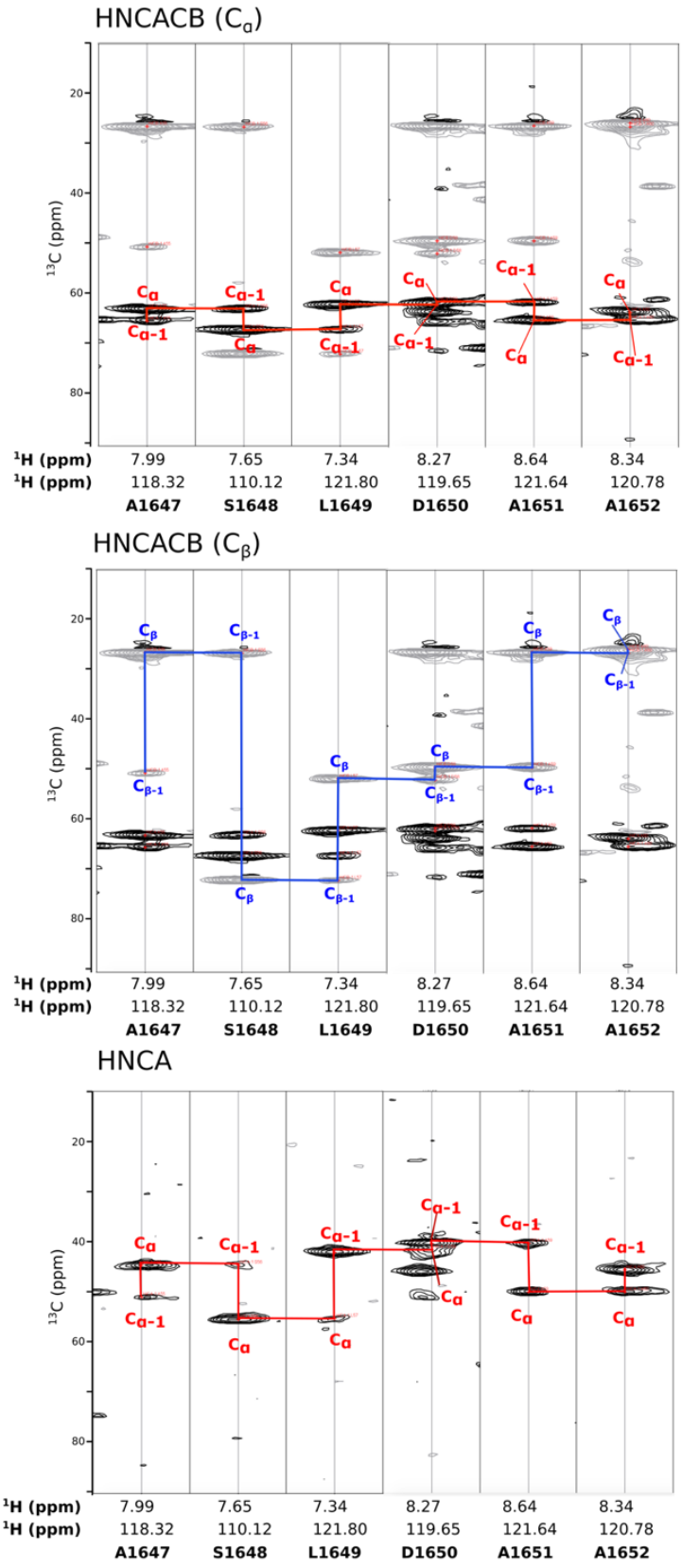


Figure 3.20 An example of an assignment and sequential walk using experiments HNCACB and HNCA.

The assignment of the backbone residues allowed a prediction of the secondary structure elements using the chemical shift index method[124]. The programme TALOS+ can be used to predict the dihedral angles which in turn are used to predict secondary structure elements. TALOS-N is a software package for the prediction of the protein backbone and sidechain torsion angles from NMR chemical shifts. TALOS-N uses secondary chemical shift alongside sequence information to derive predictions for the backbone angles ϕ and ψ and sidechain angles χ^1 with a measure of the uncertainties in each of these predictions, and also the secondary structure. It has been shown that secondary chemical shifts for C_α and C_β of a given residue correlate closely with its ϕ and ψ torsion angles, which therefore permits TALOS to perform this analysis. When given an input file of the amino acid sequence and chemical shifts for nuclei within the residues, TALOS will analyse the sequence in triplets and search a database for triplets of adjacent residues with chemical shifts and a sequence similar to those inputs to attempt to find the best match to the current triplet of interest. The database contains backbone torsion angles of proteins solved using crystal structures to a resolution of ≤2.2˚A. TALOS will then select the 10 triplets with the closest matches to the input sequence for each triplet in the query sequence. If the 10 triplets exhibit similar ϕ and ψ backbone angles, their averages can be used as angular restraints for the sequence under investigation. In addition to giving important structural information, backbone chemical shifts also provide useful insights into the flexibility of the backbone in the form of order parameters, S^2[124].

Figure 3.21 shows the secondary structure prediction per residue for *apo*-Bamb_5919 ACP using TALOS+. There was a good agreement with the parts of the homology model (Figure 3.1) that were predicted as helical. However, two residues were predicted as β-helical, which might result from a specific conformation of the two residues within the helix or misassignment. Furthermore, high S^2 values (~0.9) (Figure 3.21A) correlate well with the predicted secondary structure. Lower predicted S^2 values (~0.6– 0.8) aligned with flexible loop regions and unstructured tails. However, residues preceding helix 1, which were not predicted as helical in the homology model, were categorised as helical by TALOS+. At the same time, S^2 values of that region indicated lower flexibility. This implied that these residues might also be a part of helix 1.

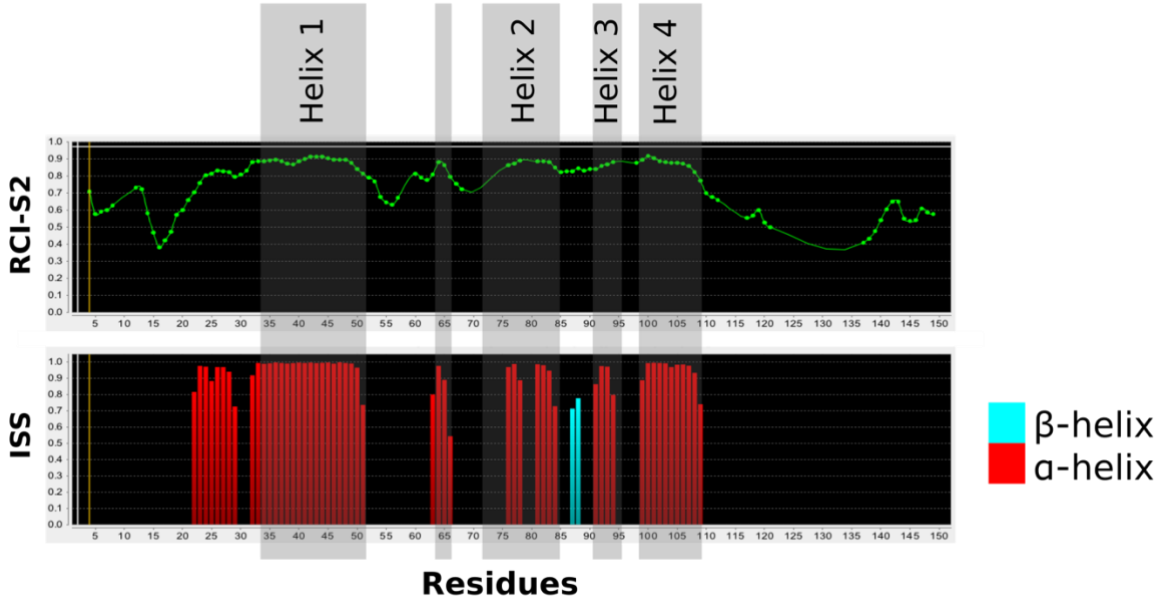


Figure 3.21 Secondary structure prediction of the assigned Bamb_5919 ACP using TALOS+. The analysis comprises predicted S^2 values (RCI-S2) and predicted secondary structure (ISS).

3.3.3 Bamb_5919 ACP protein backbone relaxation

In addition to providing detailed information about the internal dynamics of the *apo*-Bamb_5919 ACP backbone, ^{15}N relaxation experiments were recorded to help the backbone assignment of some of the residues that were complicated to assign. Thus, the relaxation analysis was used to estimate whether these residues were likely to be located in flexible or rigid parts of the protein. Also, the relaxation was used to inspect how the assigned residues behave on an NMR timescale.

Protein dynamics has increasing relevance to understanding how biological and chemical phenomena occur. Macromolecular functions are often associated with energetic transition, which is connected with structural changes and molecular flexibility[125] .

In an NMR experiment, after an RF pulse has perturbed the magnetisation from the z -axis to the xy-plane, the magnetisation does not precess there indefinitely. Relaxation is the mechanism by which the nuclear spins return to the equilibrium state. There are two key relaxation processes: spin-lattice relaxation (also known as longitudinal relaxation, described by rate R_1), which describes the return of longitudinal magnetisation back to the equilibrium state determined by the Boltzmann distribution, and spin-spin relaxation

(also known as transverse relaxation, described by rate R_2), which describes the decay of transverse magnetisation (in the xy-plane) to zero.

Many motional processes can be studied by measuring the rate of these relaxation mechanisms because relaxation is driven by the random thermal motion of the molecules. The motion experienced by the nucleus causes fluctuating magnetic fields that interact with the spins. This drives the magnetisation to equilibrium. Such fluctuations can be caused by two mechanisms: variations in the dipolar interaction between two spins or in the chemical shift anisotropy of the spin itself. The magnitude of these types of anisotropic interactions varies due to the reorientation of molecules as a result of a dynamic process, which results in the fluctuating magnetic fields. If these fluctuations occur at the required frequencies they induce transitions between spin energy levels driving the bulk magnetisation back to equilibrium. Because these contributions to the relaxation rates originate from random motions they are said to be incoherent. The correlation time and correlation function are used to define how quickly a magnetic field randomly fluctuates and whether it is at the correct frequency to cause relaxation. The correlation time (τ_c) is defined as the average time taken for a molecule to rotate one radian away from its starting position. In solution, proteins often have the overall τ_c on the nanosecond timescale, whereas small molecules will have faster τ_c on the picosecond timescale. The reciprocal of the correlation time ($1/\tau_c$) provides a guide to the average frequency of the motion.

Spin-Lattice Relaxation (R_1)

During an NMR experiment, RF pulses cause the magnetisation to deviate from the thermal equilibrium. The return of the longitudinal magnetisation back to this equilibrium (i.e. the Boltzmann distribution) is known as spin-lattice relaxation. The time taken for this relaxation process to occur is known as the spin-lattice relaxation time (T_1), alternatively the relaxation process can be described by a rate (R_1):

$$T_1 = \frac{1}{R_1}.$$

The basic method of measuring this relaxation rate involves preparing the system in a non-equilibrium state and repeatedly recording the amplitude of the signal in a spectrum while varying time t, during which the system relaxes back to equilibrium. The relaxation rate can then be extracted from an exponential fit of signal intensity against t. The spin-lattice relaxation rate contains a large amount of information about the picosecond - nanosecond dynamics of a particular nucleus. This is because the motions that are most effective at causing R_1 are on a timescale that is close to the Larmor frequency (i.e. energy gap between the and states), this is typically in the ps - ns range.

Spin-spin Relaxation (R_2)

An RF pulse will create magnetisation in the xy-plane. The decay of this transverse magnetisation to zero at equilibrium is known as spin-spin relaxation. It is quantified by the spin-spin relaxation time (T_2) or the spin-spin relaxation rate (R_2) where:

$$T_2 = \frac{1}{R_2}.$$

The R_1 data recorded for *apo*-Bamb_5919 ACP showed a similar trend across the protein, which suggests that the majority of the assigned residues are characterized by similar dynamics on the fast ps-ns time scale (Figure 3.22). However, more significant differences are observed in the transverse relaxation rate constants (R_2). The parts of the protein that were predicted as helical from the homology model displayed shorter relaxation times suggesting more structured conformations in these regions. Several residues preceding helix 1 also displayed short relaxation times, implying that these residues might be also a part of the helix 1, as suggested by TALOS+ prediction (Figure 3.21).

Several residues showed unusual R_1 dynamics profile (T1709, S1765, A1653, H1762, S1765 and S1766). They have been either misassigned or the proper integration of their peaks was not possible.

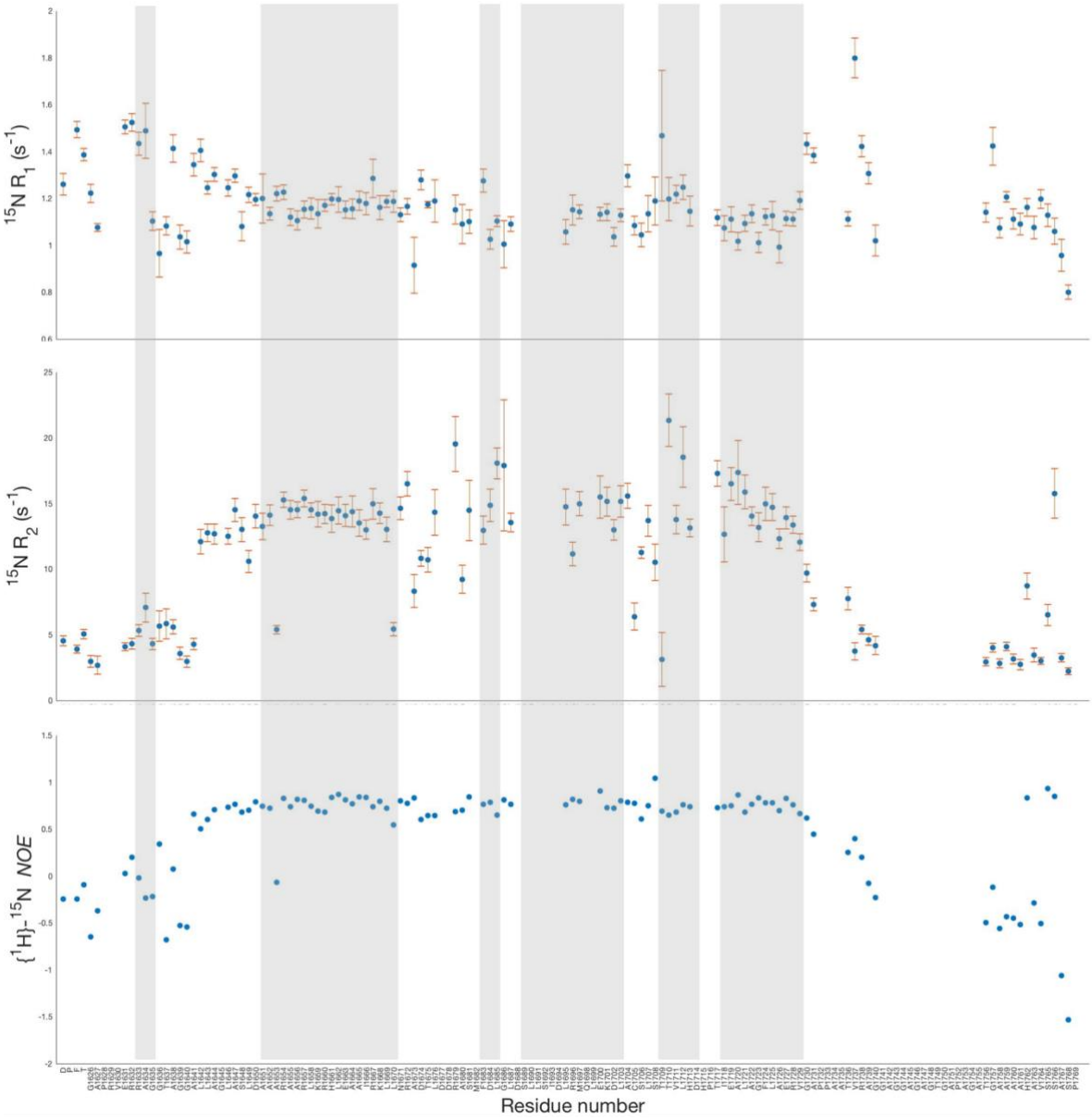


Figure 3.22 Solution-state NMR dynamics data for the apo-Bamb_5919 ACP domain. Grey strips show the location of helical elements predicted from the homology model. Data were recorded at 700 MHz and 298 K. Error bars represent the standard deviation of the measurements. The gaps represent unassigned residues.

T_1/T_2 ratio can be used as an estimate of the overall rotational correlation time (τ_c). All residues were included in the calculation that was based on:

$$\tau_c = \frac{1}{4\pi\nu_N}\sqrt{6\frac{T_1}{T_2} - 7},$$

where ν_N is the ^{15}N resonance frequency (in Hz) [126].

A value for the relaxation-derived τ_c of 7.93 ns was slightly longer than the expected value of 7.4 ns obtained using the general rule of 0.5 ns τ_c per 1 kDa molecular weight[127–129], taking into account isotopic enrichment and the presence of affinity purification tags. This deviation from the ideal value is not large enough to imply oligomerization of *apo*-Bamb_5919 ACP, but may reflect a non-spherical shape of the monomeric protein and could also arise from the presence of long unstructured termini. This was in accordance with the obtained analytical size-exclusion chromatogram that was additionally used to inspect the oligomerisation state of the *apo*-Bamb_5919 ACP in the buffer conditions used for NMR studies (Figure 3.23). Namely, the protein mass of both Bamb_5919 ACP and Bamb_5919 KS^0 calculated based on the elution volume corresponded to trimer. This was not surprising as τ_c suggested that *apo*-Bamb_5919 ACP was not perfectly globular. In addition, CD analysis of both constructs showed that 23-28% of each protein is disordered.

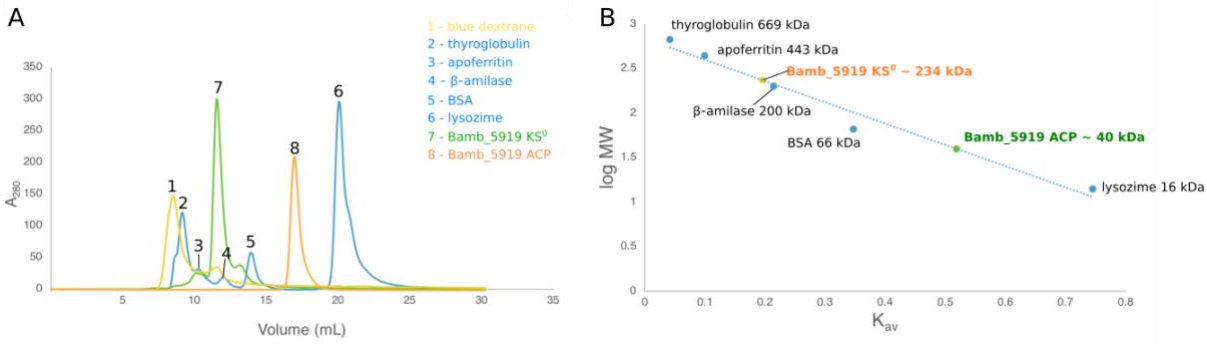


Figure 3.23 Analytical SEC analysis of Bamb_5919 ACP and Bamb_5919 KS^0.

The disordered termini are highly mobile on fast ps-ns timescale compared to the structured core, as indicated by lower heteronuclear NOEs (Figure 3.22).

3.3.4 ^{15}N-HSQC analysis of apo- to holo- conversion of Bamb_5919

To get an insight into which residues are affected by the conversion of Bamb_5919 ACP from *apo* to *holo* form, ^{1}H-^{15}N-HMQC spectra were recorded for each form (Figure 3.24A). The assignments of the backbone residues were transferred to the ^{15}N-HMQC spectrum recorded for *holo* sample. The differences in peak shifts between two protein forms were

analysed as chemical shift perturbations (described in the next paragraph) (Figure 3.24 B).

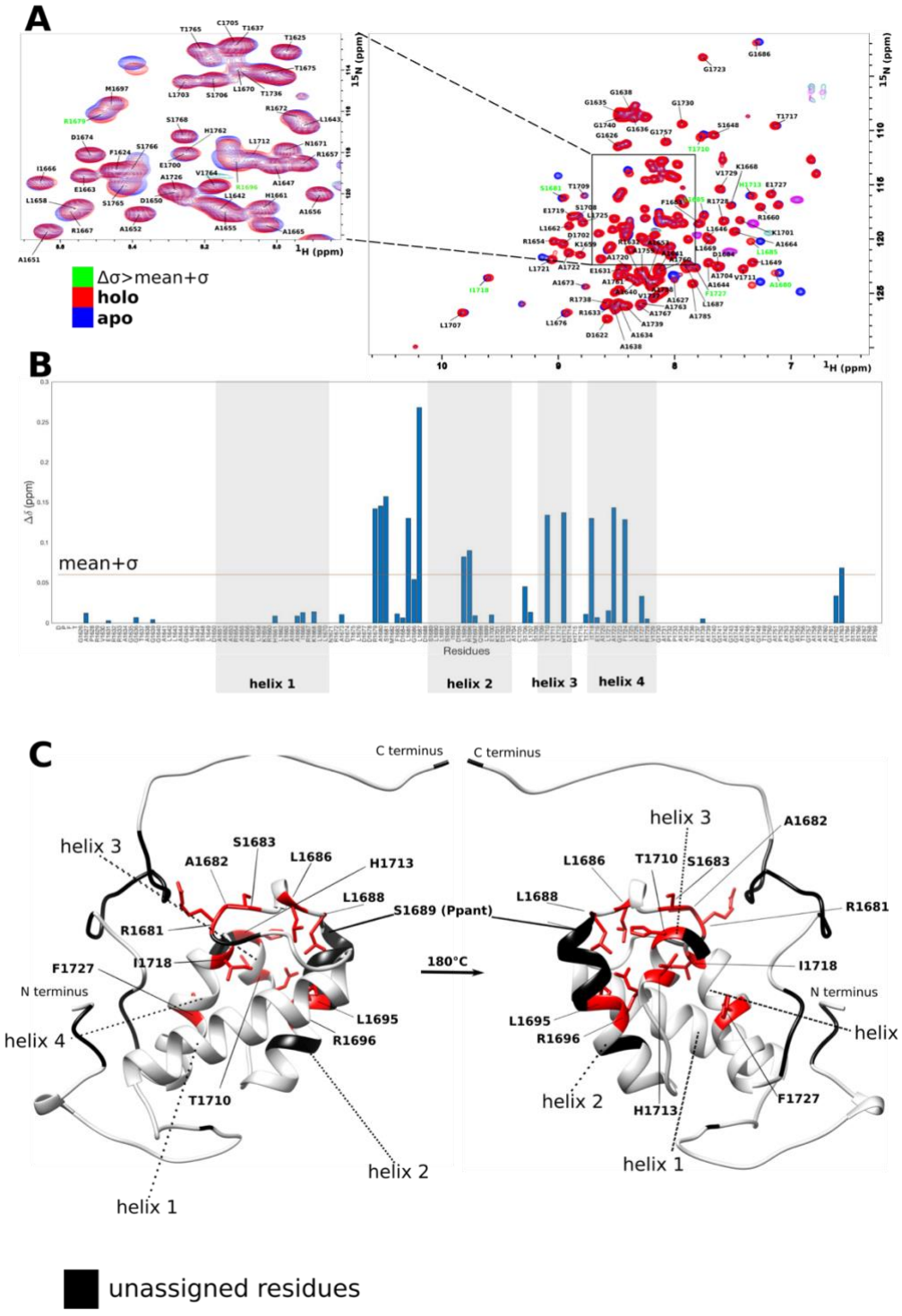


Figure 3.24 Chemical shift difference between apo and holo form of Bamb_5919 ACP. A) The overlap between ^{15}N-HMQC spectra recorded for apo and holo forms. B) Average

chemical shift differences, $\Delta\delta_{av}$, shown between holo and the apo form, C) The residues with $\Delta\delta_{av}$ above the average in B) are shown mapped on the homology model of Bamb_5919 ACP.

The chemical shift perturbations observed on conversion could be interpreted as evidence that the Ppant arm mostly affects the conformations of peaks that are oriented towards the turn between the first two helices. Several residues helices 3 and 4 were also affected (Figure 3.24 B). A similar pattern was observed in ACPs from other systems[124,130].

Investigating Bamb_5919 ACP:Bamb_5919 KS^0 interface by solution-state NMR titrations

NMR titrations were used to map the binding interface between *apo* and *holo* forms Bamb_5919 ACP and Bamb_5919 KS^0 domain. Due to its size, it was not possible to obtain the backbone assignment of Bamb_5919 KS^0. For that reason, it was possible to monitor the effects of binding only on Bamb_5919 ACP.

Generally, NMR can be used to study protein-ligand and protein-protein interactions in several ways. Typically, an NMR study of a protein-protein interaction starts with titration experiments, in which an isotope labelled protein is observed (for example using the ^{1}H-^{15}N-HSQC) in the presence of increasing concentrations of an unlabelled binding partner. If the two proteins interact, the labelled protein will undergo chemical shift perturbations (CSPs) caused by changes in the chemical environment around residues that are affected during the interaction. CSPs can be used to map the binding interface which is known as chemical shift mapping[124]. During a titration experiment, the proteins exchange between free and bound states at a rate which depends in part upon the binding affinity and the sample temperature. The exchange between free and bound states occurs at different rates, which can be categorized into three exchange regimes: fast, intermediate and slow. The absolute rate of exchange is not relevant for analysis, but the rate of exchange (k_{ex}) relative to the difference in the chemical shift (or frequency, $\Delta\omega$) between a residue in the free and bound states (Figure 3.26). When $k_{ex} > \Delta\omega$, the exchange is said to be in the fast regime, when $k_{ex} < \Delta\omega$ the exchange is slow and when $k_{ex} \approx \Delta\omega$ the exchange is intermediate. For a residue exchanging within the fast regime, a single peak is observed, which represents a population-weighted average between the free and bound states. The

fast exchange typically occurs for interactions with a Kd > 3 µM and is recognised in a two-dimensional spectrum such as the ^{15}N-HSQC when a peak appears to shift from one position to another with increasing concentrations of ligand. For a residue exchanging in the slow regime, two peaks are observed. One peak remains at the original position, whilst the other one emerges at the new position. The intensity of the original peak decreases as the intensity of the new peak increases. This happens because the population of protein in the free state decreases and population in the bound state increases. Intermediate exchange is characterised by chemical shift changes that are accompanied by line broadening, which can result in the complete disappearance of a peak. For most systems, exchange on the millisecond timescale will fall into the intermediate limit.

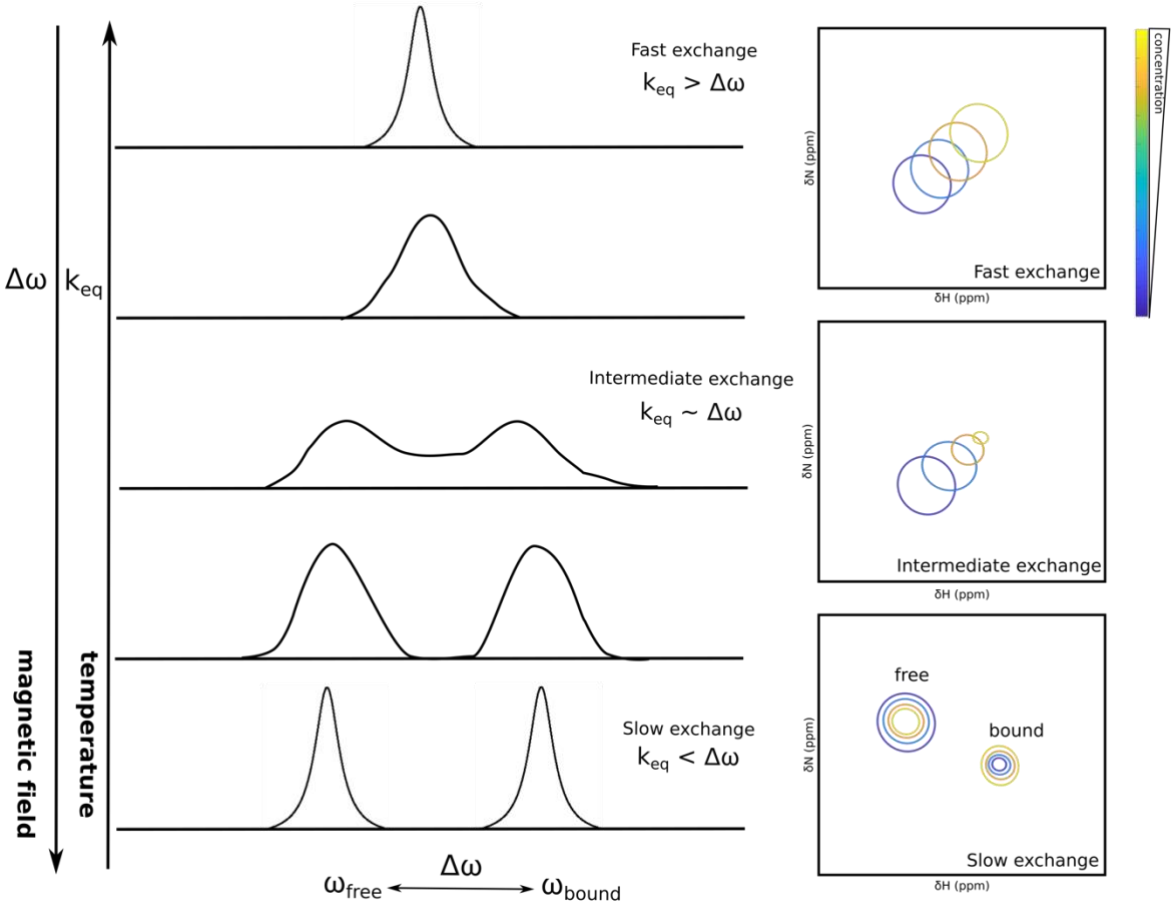


Figure 3.25 Exchange regimes in NMR. The diagram illustrates the slow, intermediate and fast exchange of a single resonance between two different conformational states (ω_{free} and ω_{bound}). In the slow regime, two peaks are observed. In the fast regime, a single peak

is observed at an average position. K_{ex} increases with temperature and Δω increases with a magnetic field. Examples are shown of each regime in a ^{1}H-^{15}N HSQC.

NMR titrations were used to analyse the interaction between both apo and holo-Bamb_5919 ACP to inspect whether the presence of a covalently bound prosthetic group affects the interaction profile.

Upon addition of Bamb_5919 KS^0, no peak shifts were observed. The intensity of several peaks in the spectrum was decreasing along the ligand concentration gradient (Figure 3.27).

This indicated that the interaction was following a slow exchange regime. However, no new peaks were forming in the spectrum. This was due to an increase in the correlation time of ^{15}N-Bamb_5919 ACP upon complexation with Bamb_5919 KS^0 so the resulting complex was ~ 90 kDa compared to 15 kDa of the unbound Bamb_5919 ACP. The rate of transverse relaxation T_2 is proportional to the correlation time so the signal intensity is expected to decrease with slower tumbling.

The peak intensity decay rates can be used to calculate K_d of individual residues. However, these values should not be treated as exact values, such as K_d values obtained from BLI or ITC. Still, their analysis was used to inspect whether the overall pattern of the backbone K_d values would be similar between apo and holo datasets and to monitor which peaks are the most affected by the presence of Bamb_5919 KS^0. The K_d values that were obtained for four replicates of *apo* sample were relatively consistent across parts of the protein that were predicted as helical (K_d = ~ 60 μM)(Figure 3.28). Contrarily, the K_d values of the residues likely to be in the flexible part of the protein were inconsistent and higher. These peaks were mostly located in the region of HMQC with a significant level of signal overlap and were thus not possible to properly integrate. Their fits were not good in some cases (Supplementary). The relative K_d values obtained for three replicates of holo samples were higher compared to the K_d pattern of apo samples.

Altogether, this suggested the stronger binding of *holo*-Bamb_5919 ACP compared to apo-Bamb_5919 which was consistent with the BLI data (Figure 3.9), but also confirmed that unstructured termini seem not to participate in the interaction with Bamb_5919 KS^0.

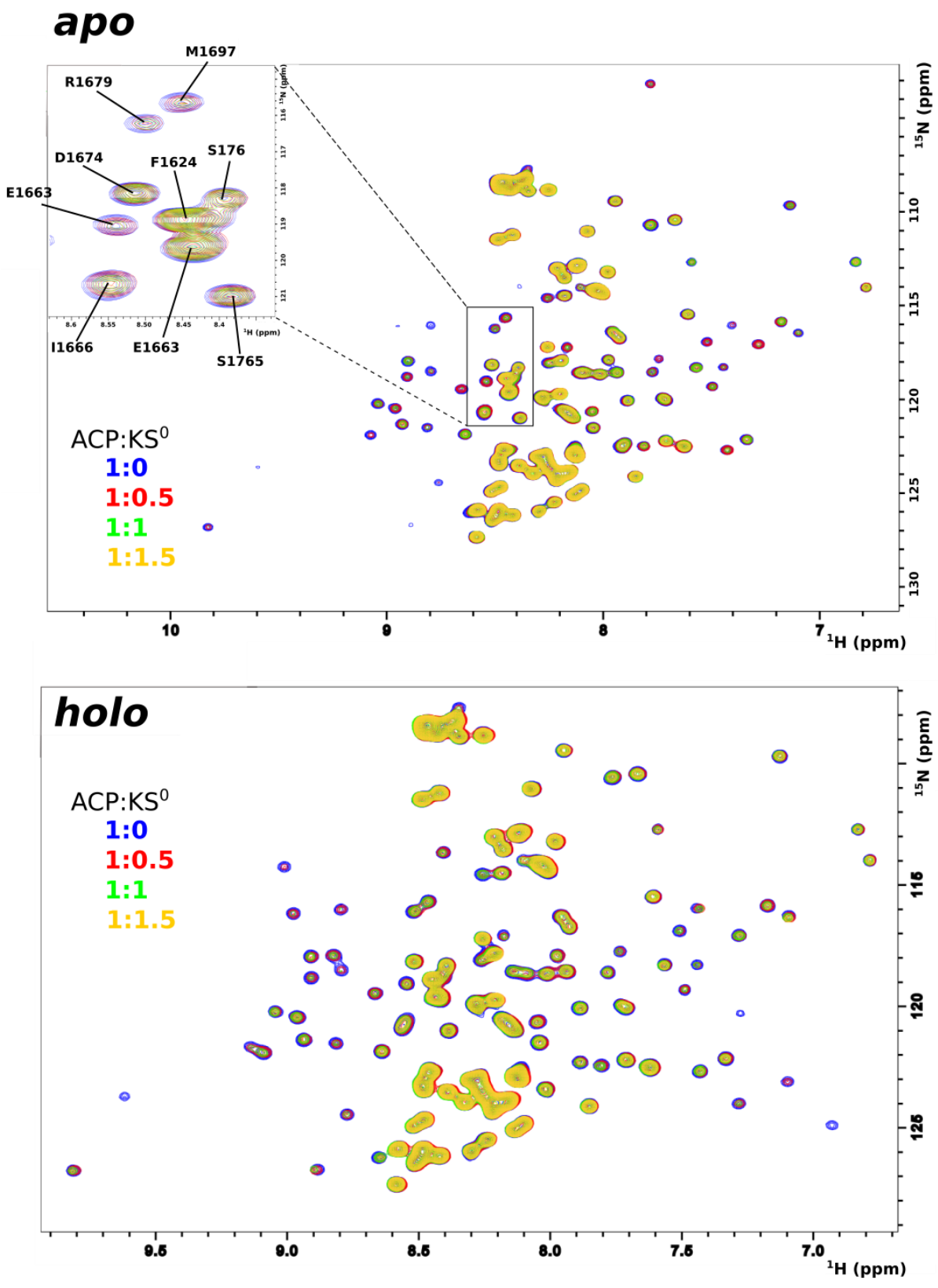


Figure 3.26 ^{1}H-^{15}N HMQC titrations of apo- and holo-Bamb_5919 ACP with the increasing concentrations of Bamb_5919 KS^{0}. The addition of ligand caused some peaks to decrease in intensities. The intensities of other peaks, the majority of which was located in the part of the spectrum related to unfolded parts of the protein, were not affected.

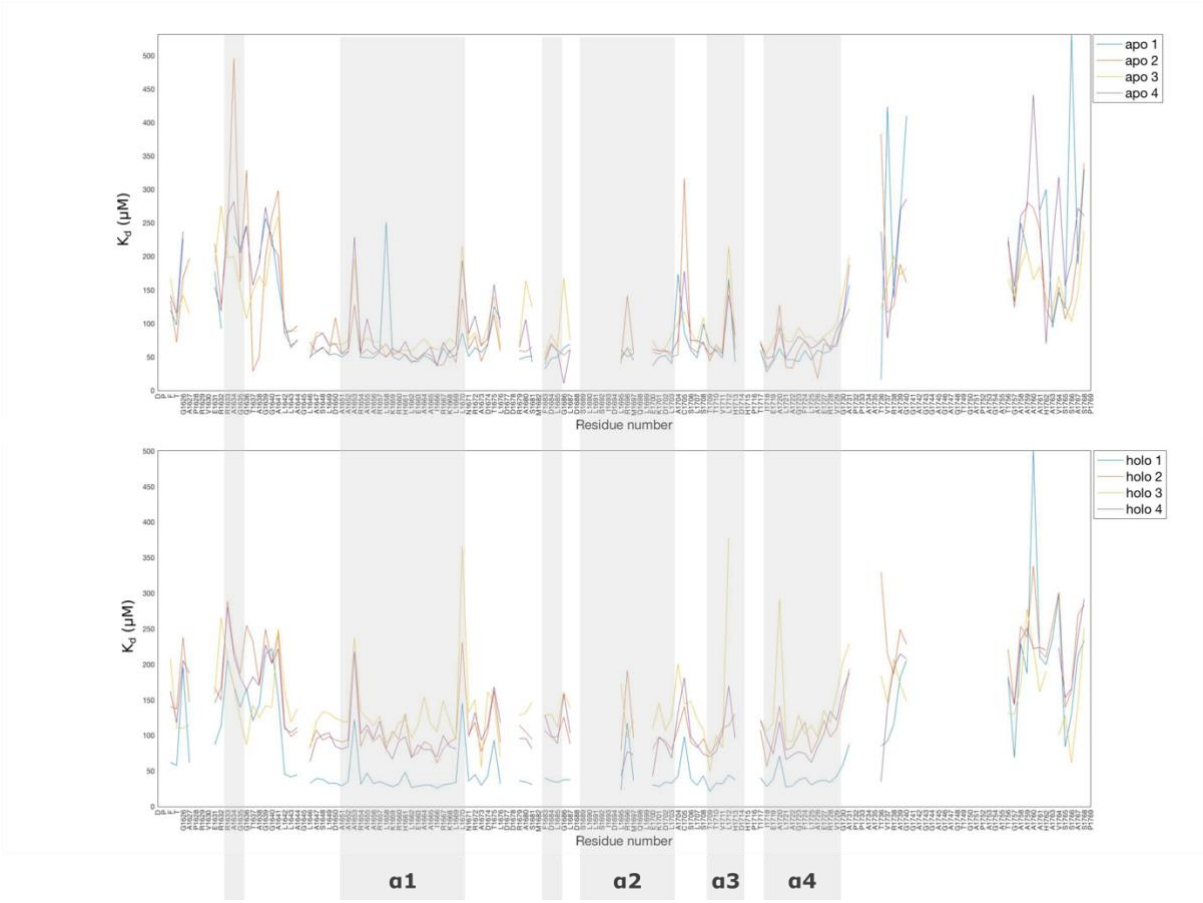


Figure 3.27 Local apparent Kd values obtained from solution-state NMR titrations. The plots represent all replicate experiments recorded for titrations of A) apo-Bamb_5919 ACP and B) holo-Bamb_5919 ACP with Bamb_5919 KS^0.

Next, to distinguish which individual residues experience the highest peak intensity decay upon binding, K_a was calculated based on:

$$K_a = \frac{1}{K_d},$$

and plotted in Figure 3.26. To inspect the location of the residues that had the highest decay rate (higher than the mean + standard deviation, as described in Chapter 2), they were mapped onto the homology model of apo-Bamb_5919 ACP for both *apo* and *holo* Bamb_5919 ACP (Figures 3.29 and 3.30).

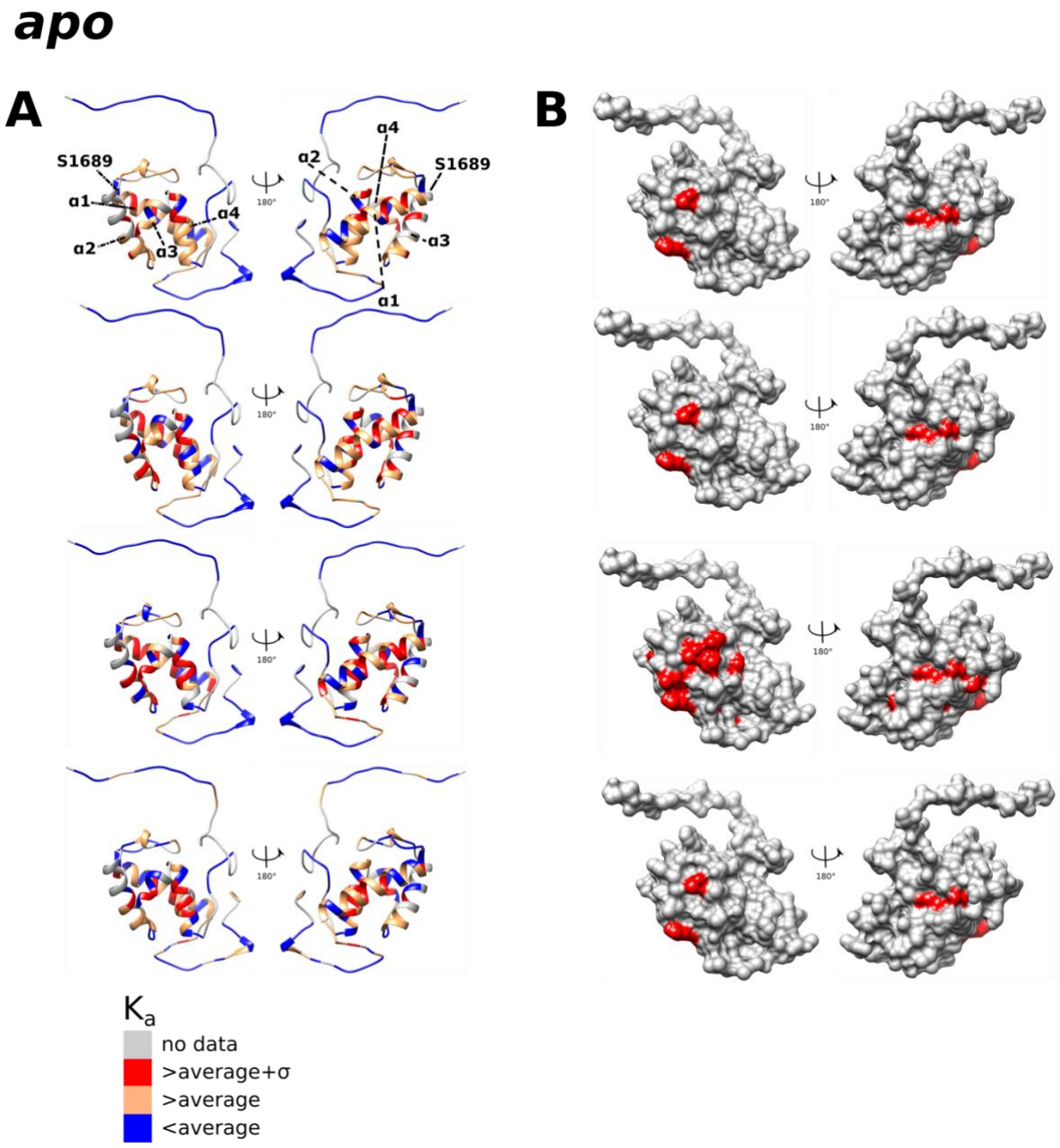


Figure 3.28 Local apparent K_a obtained from NMR titrations of apo-Bamb_5919 with Bamb_5919 KS^0 projected onto the structure of the homology model of holo-Bamb_5919 ACP (Figure 3.1). Each dataset is presented. Majority of residues with a higher than average local apparent K_a were located in the parts of the protein that were predicted to be structured (helical).

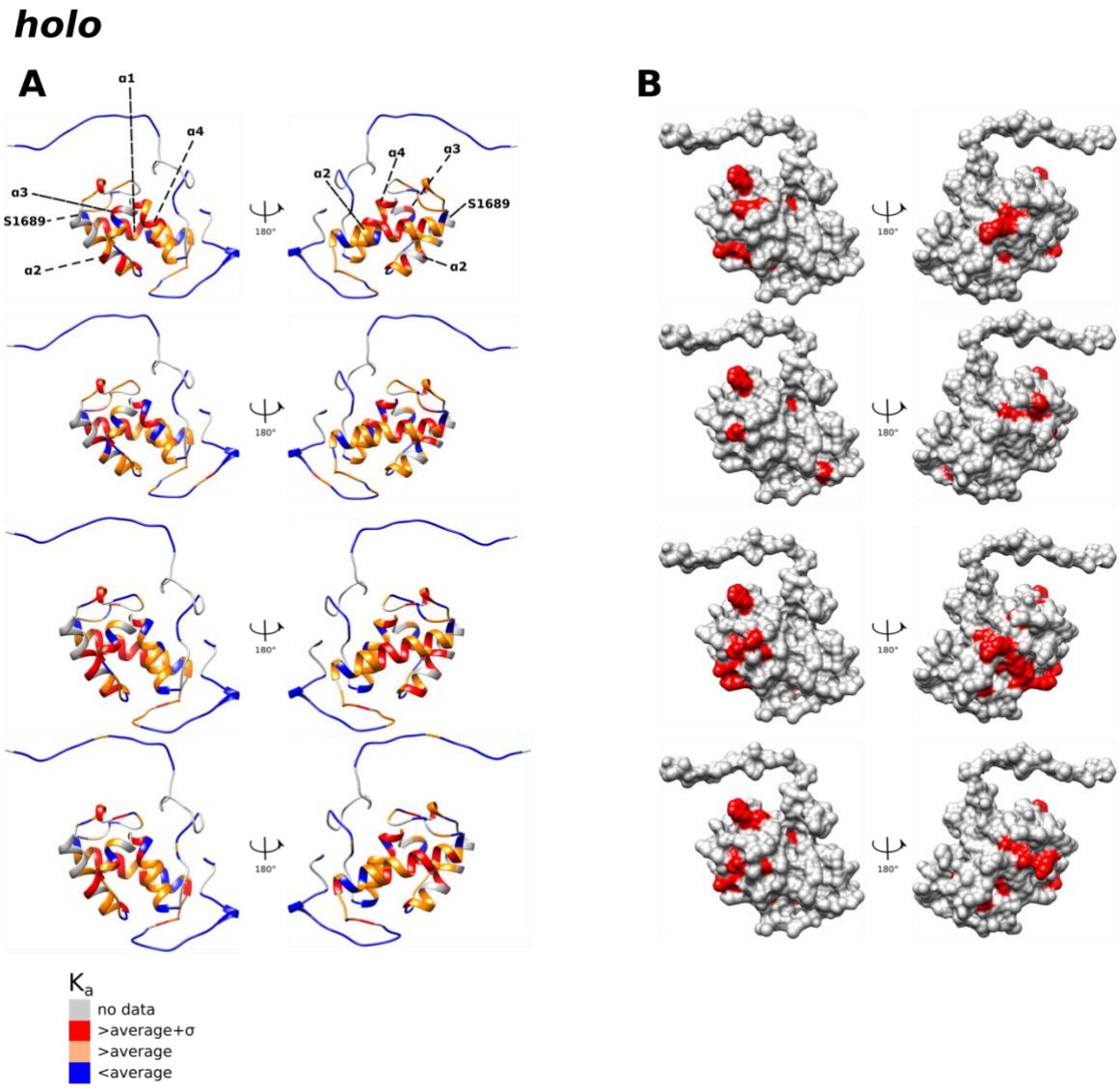


Figure 3.29 Local apparent K_a obtained from NMR titrations of holo-Bamb_5919 with Bamb_5919 KS^0 projected onto the the homology model of apo-Bamb_5919 ACP (Figure 3.1). The results obtained for four replicate datasets are shown separately.

Based on the data presented in Figures 3.29 and 3.30., the residues that are the most affected by the presence of Bamb_5919 KS^0 were located in the core of the protein and do not seem to form a surface patch. This indicated conformational change in Bamb_5919 ACP upon binding. Furthermore, to compare the patterns of affected residues in all apo and holo datasets, Figure 3.30 was created. Again, the analysis of K_a values demonstrated that N and C-termini probably do not interact with KS^0. Thus, their protein sequence is not shown in the figure.

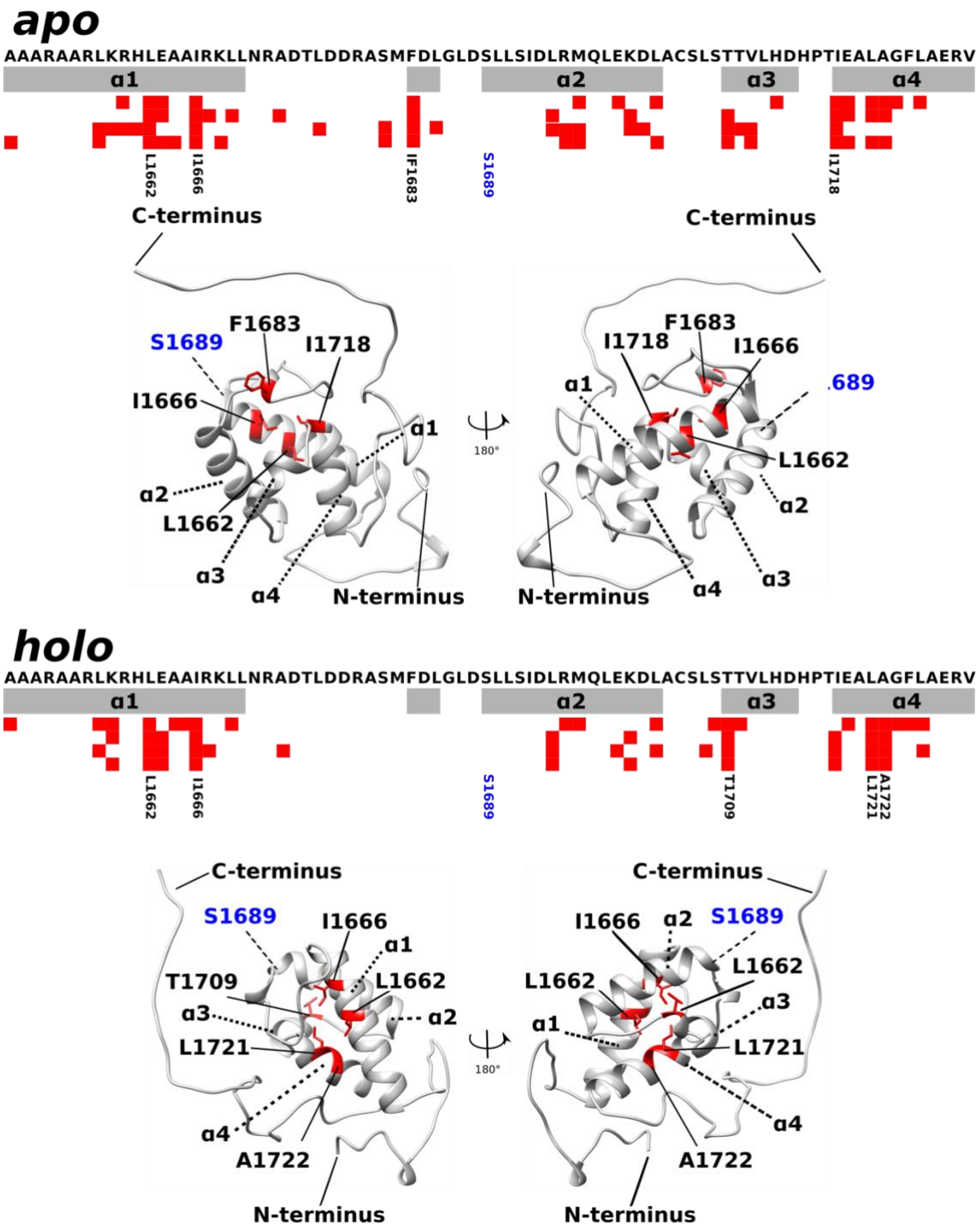


Figure 3.30 Higher than average local apparent Ka projected onto homology model of Bamb_5919 ACP. The protein sequence, not including termini, is also shown. The red squares represent the residues with the highest intensity decay upon addition of ligand in all datasets. Serine accepting Ppant arm is indicated.

Even though all the helices were affected to some extent in both apo and holo proteins, several hydrophobic residues were the most affected in all datasets. These residues were L1662 and I1666 that are located in helix 1, as well as the residues in helix 4 (I1718 in apo and L1721 and I1722 in holo sample). Thus, the Bamb_5919 ACP hydrophobic pocket might be crucial for recognition of Bamb_5919 KS^0 through hydrophobic contacts. The N-terminal part of helix II could not be properly assigned so it was not possible to monitor the effect of the presence of Bamb_5919 KS^0 on that part of the protein. However, there were no significantly affected peaks that were unassigned, which suggests that the unassigned part of helix 2 might not be significantly affected by the addition of Bamb_5919 KS^0.

However, several studies have suggested that negatively charged, solvent-exposed amino acids within helix II ACPs may have important roles in protein-protein interactions and act as a recognition motif for the interaction of different types of ACPs with their partner domains[52,131–136]. For example, the docking model of type II *E. coli* ACP interacting with KS III showed that this interaction is mediated by a hydrophobic and positively charged surface patch on KS III which docks with a partly hydrophobic and negatively charged surface on the helix II of ACP[133]. The study also highlighted that the Glu, which is located just above Ser in ACP forms a crucial ionic interaction adjacent to the KS active site entrance.

Furthermore, photo-crosslinking assay and mutagenesis experiments of the same system suggested that the charged residues in helix II are crucial for its interaction with the ketosynthase[137]. The recent study of the same system highlighted that Asp at the C-terminal and Glu at the top N-terminal part of the helix II are particularly important for ACP-KS interaction[135]. The molecular dynamics experiments in the same study also showed that ACP in the apo state loses its ability to maintain several of the salt bridge contacts observed in the crystal structure. However, upon being activated to its holo state, salt bridge contact times increase on the bottom of helix II. Interestingly, upon loading of the Ppant arm with a substrate, ACP helix III begun to make new interactions with the ketosynthase that were not found in the crystal structure. A similar effect was observed in *holo*-Bamb_5919 ACP where T1709 in helix III seems to be more affected by the presence of Bamb_5919 KS^0 in comparison with *apo*-Bamb_5919 where the residues in helix III did not seem to undergo a local conformational change.

For type I PKS, which is the same type of PKSs as enacyloxin PKS, studies that were focused on the protein recognition features of ACP domains also identified helix II as potentially important in KS–ACP interactions[138,91]. It was even proposed that several charged residues on helix II might enable ACPs to differentiate between their partner KS domains[139].

However, in DEBS type 1 PKS (a prototypical type of PKSs described in Chapter 1), in addition to helix II, the N-terminal part of helix I in ACP2 from the third module seems to be the primary specificity determinant during intermodular chain translocation[140] (Figure 3.31). The NMR titrations between ACP2 and the didomain comprised of the interacting KS and AT, KS3-AT2, also pointed that the residues in helix 2, in addition to the loop connecting helix 2 and 3, was the most affected by the binding to the didomain[141]. Moreover, the same study proposes that during intermodular chain translocation ACP2 and KS3 engage each other via a two-step mechanism. In the first step, acyl-ACP2 docks onto module 3 via protein–protein interactions centred on helix I of the ACP as well as the short peptide sequence that lies beyond the C-terminal end of the core ACP domain. In the second step, insertion of the acyl chain into the KS active site leads to further tightening of the entire complex and the transfer of the growing polyketide chain.

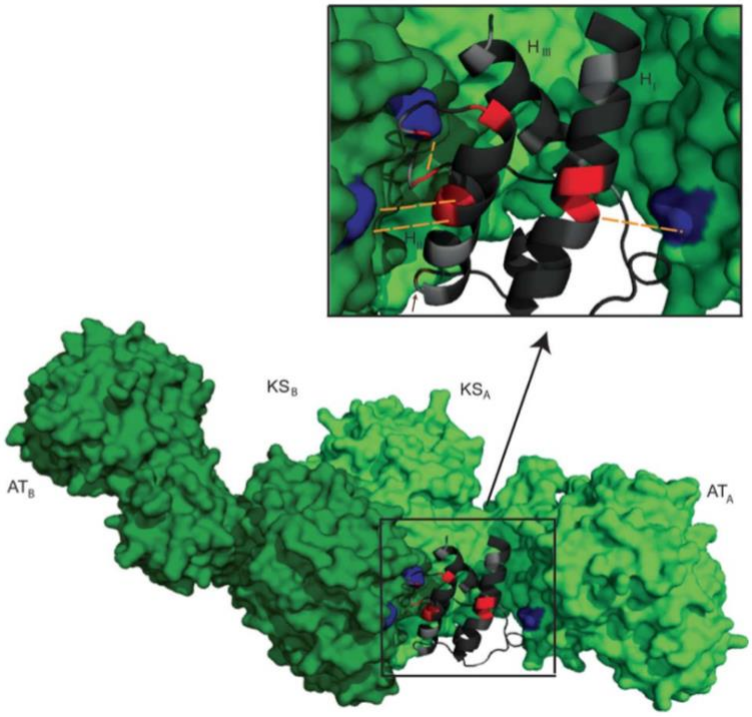


Figure 3.31 Proposed docking model for ACP2 (dark grey) onto [KS3][AT3] (green) during the chain transfer mode of interaction in DEBS[141]. Apo-ACP2 residues with solvent-exposed side chains whose backbone HSQC peaks significantly decreased in intensity upon addition of [KS3][AT3] are highlighted in red. These residues are L24, R26, N62, R61, T68, and T75. Salt bridges and hydrogen-bonding interactions between the ACP2 residue (red) and [KS3][AT3] partner residue (blue) are highlighted in orange. Unassigned ACP2 residues are highlighted light grey. The figure was taken Charkoudian LK, Liu CW, Capone S, et al., 2011.

Overall, solution-state NMR provided information about the residues of Bamb_5919 ACP that were the most likely to be involved in the interaction with Bamb_5919 KS^0, i.e. ^{15}N-Bamb_5919 KS^0. However, the same solution-state NMR approach could not be used to characterise the binding interface on Bamb_5919 KS^0. Attempts have been made previously to crystallise KS^0 by Dr. J.C.Milligan, a member of collaborating prof. Tsai group[142]. However, the diffraction quality of KS^0 crystals was found to be relatively poor. Combining datasets from several different crystals enabled processing to a resolution of 3.05 Å.. Also, It was not possible to obtain phases by using molecular replacement.

Therefore, as the X-ray crystal structure of enacyloxin module 10 KS^0 is unknown, it was not possible to determine the interacting residue partners in Bamb_5919 KS^0.

3.3.5 The solid-state NMR analysis of the Bamb_5919 ACP:Bamb_5919 KS^0 complex

Solution-state NMR analysis of proteins and protein-protein complexes is limited by the broadening of resonances at higher molecular weights and the losses of spectral resolution and sensitivity. For that reason, it is not possible to observe the signals originating from Bamb_5919 ACP backbone when the complex with Bamb_5919 KS^0 is formed. It is only possible to determine which residues are likely involved in the interaction with the ligand, which was attempted in the previous section. Thus, the information about the shifts and conformation of these residues in the complex cannot be determined by solution-state NMR.

Solid-state NMR spectroscopy has emerged as one of the few techniques that can provide atomic-level structural information for these types of systems. Solid-state NMR takes advantage of the fact that the line widths in solids do not depend on the size of proteins as they do in solution. Hence, by using solid-state NMR to study complex between Bamb_5919 ACP and Bamb_5919 KS^0 it might be possible to get an NMR fingerprint of their interface in the complex.

The ability to apply solid-state NMR to study large complexes became possible only recently thanks to the progress facilitated by advances in sample preparation and NMR methodology. Solid-state NMR was applied to study for example amyloid fibrils[143], membrane proteins[144] and complex molecular assemblies[145]. With regards to sample preparation, an important recent development was the introduction of FROSTY (freezing rotational diffusion of protein solutions at low temperature and high viscosity) sedimentation by ultracentrifugation as a method to prepare homogenous protein solid-state NMR samples that yield narrow resonances suitable for detailed quantitative studies[146]. This approach has enabled the solid-state NMR study of a very large, 1.1 MDa, highly symmetric $\alpha_7\beta_7\beta_7\alpha_7$-11S multimeric complexes from the 20S proteasome. So far, the only asymmetric system studied by this approach has been the complex between Bamb_5919 PCP and Bamb_5915 C, which was analysed by Dr. A. Gallo and Dr. S. Kosol, members of Lewandowski group[147]. FROSTY was also used in this PhD project to prepare the sediments between Bamb_5919 ACP and Bamb_5919 KS^0.

For solid-state NMR investigation, the protein samples have to be properly hydrated, which is crucial for their activity, folding and dynamics[148–153]. On the other hand, sample

dehydration can reduce the spectral resolution. Sedimentation by FROSTY can be also used to immobilise the proteins by centrifugation while retaining the homogeneous local environment, which is useful if the sample cannot be crystallised. The sample has to be evenly packed into the rotor to maximise the sample quantity. Hydrated proteins are semi-solids and typically stick to the tools used to pack dry samples, while simultaneously being too viscous to be transferred using a pipette. Moreover, sedimented proteins must be transferred to the rotor with as little waste of the costly sample as possible, while also not changing the hydration of the protein. In the Lewandowski lab, proteins are usually packed into 0.7 and 0.8 mm rotors manually by using a spatula and into 1.3 mm rotors using a benchtop centrifuge.

Solid vs. solution-state NMR

The NMR interactions that affect the Larmor frequency of a nucleus are the Zeeman interaction (the interaction with an external magnetic field $B_{0)}$[154], dipolar coupling and chemical shift anisotropy. The precise Larmor frequency of the nucleus in a magnetic field does not only depend upon the type of nucleus (i.e. gyromagnetic ratio), but also the surrounding environment. The chemical shift or chemical shielding, like the Larmor frequency, is proportional to the external magnetic field. Electrons are not spherically distributed around a nucleus, therefore shielding from the electrons will be anisotropic. This means that the orientation of the nucleus with respect to the B_0 field will affect the chemical shift, which is known as chemical shift anisotropy (CSA). In the solution state, where the molecules are rapidly tumbling, the anisotropy is averaged out and just the isotropic chemical shift is observed. But in the solid-state CSA is an important factor to consider because the molecules will be in different orientations with respect to the B_0 field and each will produce a slightly different chemical shift. The resulting peak will be a superposition of all orientations and have a broad shape. In the solid state the effects of J-coupling are rarely resolved, because their influence usually has the smallest magnitude. Dipolar coupling is a through-space interaction between a pair of dipole moments. Like CSA, this interaction is also anisotropic: each orientation will give a different splitting, which causes the broadening of an overall NMR signal. The dipolar

interactions are distance-dependent so a spin will have the strongest dipolar interactions with those closest in space.

Magic angle spinning

Magic angle spinning (MAS) is applied in solid-state NMR experiments to remove the line-broadening effects of CSA and dipolar couplings, which have a common dependence on orientation:

$$\frac{3cos^2-1}{2},$$

Where θ is the angle between the B_0 field and the CSA principal axes. The rotation of a solid sample at 54.74° to the B_0 field reduces Equation 3.10 to zero. This angle has been named the magic angle (Figure 3.33). To average CSA or dipolar interactions, the frequency of the magic angle spinning (MAS) has to be larger (faster) than the size of the interaction. For example, homonuclear dipolar couplings are more difficult to average out compared to heteronuclear dipolar couplings. MAS of up to 60 kHz is typically used in protein solid-state NMR experiments, however more recent advances in NMR equipment now allow for MAS up to 150 kHz (Figure 3.33).

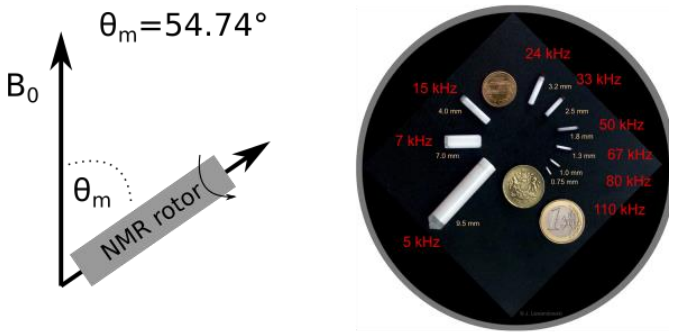


Figure 3.32 Schematics showing the MAS rotor oriented at the magic angle with respect to the external magnetic field, B0. Also shown are rotors used in MAS NMR experiments, showing outer diameters and maximum MAS rates.

Cross polarisation

The signal from a typical protein solid-state NMR experiment is low so the same experiment has to be repeated many times to increase the signal-to-noise ratio. Cross polarisation (CP) is a technique used to provide information on different nuclei in a single experiment. CP is also used to increase the signal-to-noise ratio and decrease experimental time. CP involves the transfer of magnetisation from the nuclei of an abundant isotope (e.g. ^{1}H) to a less abundant isotope, such as ^{13}C or ^{15}N. The achieved signal enhancement depends on the experimental setup, including the optimisation of the pulse powers and lengths. After running the pulse sequence the magnetisation needs return to equilibrium before repeating the experiment, which is T_1 dependent. A recycle delay, which is at the end of a pulse sequence, allows time for T_1 relaxation. This strong pulse called spin-lock ensures that the magnetisation remains along the axis as the RF field strength is greater than any offset that would normally cause deviation from this axis. The spin-lock pulse is applied along the same axis as the magnetisation vector.

^{1}H-X CP starts with a ^{1}H 90° pulse which brings the ^{1}H spins from the z-axis into the xy-plane. Spin-lock pulses are applied simultaneously on both the ^{1}H nuclei and the X nuclei for a length of time termed the contact time, which usually lasts a few ms. Magnetisation is transferred from the ^{1}H nuclei to the X nuclei during the contact time (Figure 3.34). The detected signal originates from X. ^{1}H decoupling is applied during the acquisition time.

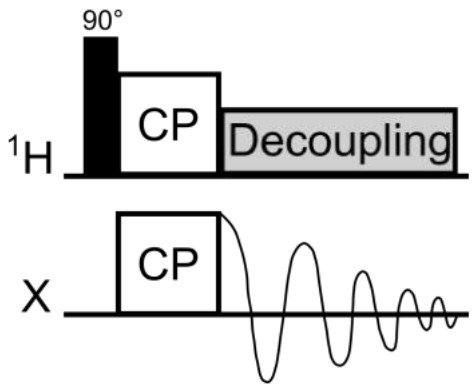


Figure 3.33 ^{1}H-X CP pulse sequence.

The polarisation transfer occurs when the Hartmann–Hahn matching condition is satisfied, which is when the nutation frequencies of the nuclei ^{1}H and X are matched by the choice of appropriate RF field strengths. Fulfilling the Hartmann-Hahn condition will cause both the ^{1}H and X nuclei to have the same energy gap between their spin states,

so that the ^{1}H-X dipolar coupling allows redistribution of energy between them and ultimately the transfer of magnetisation from ^{1}H to X.

Sedimentation of *holo*-Bamb_5919 ACP:Bamb_5919 KS^{0} complex

The initial sedimentation conditions were selected based on the calculations made by Dr. Rebecca Stevens from Lewandowski lab (Figures 3.34 and 3.35). However, the sample packing was performed manually, without the use of packing tools (MLS-50 and MLA-150). The initial sedimentation speed was thus set to 700 000 g for 24 hours.

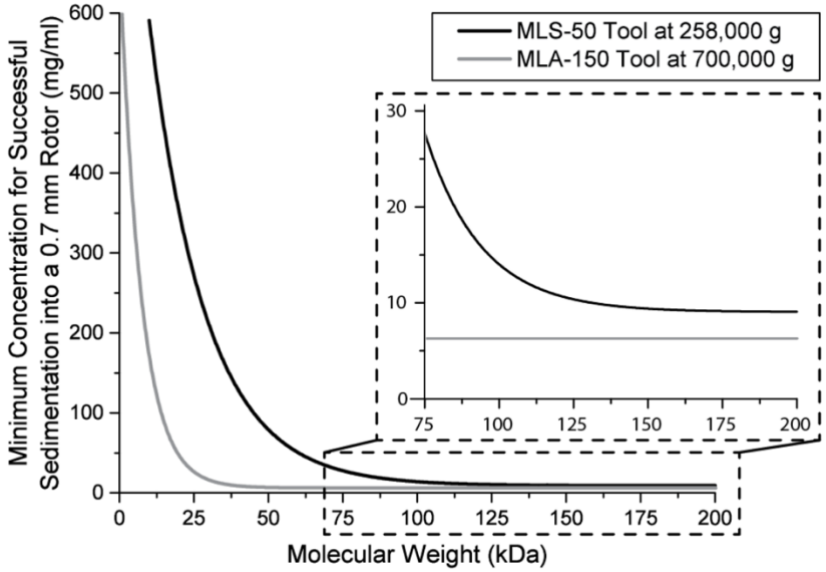


Figure 3.34 Calculated conditions required to fully sediment a protein into a 0.7 mm rotor using the MLS-50 tool at 258,000 x g (black) and the MLA-150 tool at 700,000 x g (grey). The graph indicates the minimum concentration of protein needed for a range of molecular weights. The figure was taken from dr. Stevens's PhD thesis.

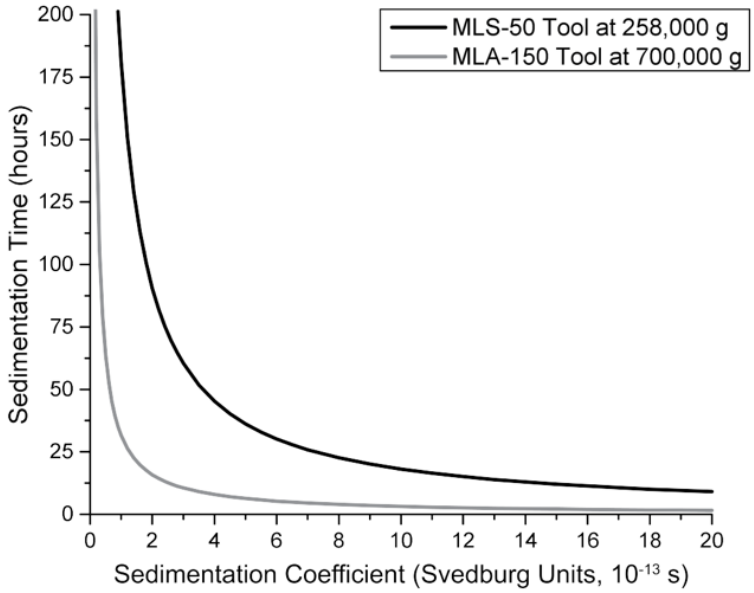


Figure 3.35 The time required for complete sedimentation using the MLS-50 and MLA. Figure 3.21: The time required for complete sedimentation using the MLS-50 and MLA-150 packing tools at 258,000 x g and 700,000 x g, respectively. The figure was taken from dr. Stevens's PhD thesis.

Compared to solution-state NMR, produces a much weaker signal. The amount of ^{15}N labelled protein in the sediment and being subsequently packed into the rotor is much smaller compared to the amount of protein inside solution-state NMR tube. For that reason, every effort was made to ensure the highest possible concentration of ^{15}N-Bamb_5919 ACP:Bamb_5919 KS^0 complex inside the rotor.

Only holo-Bamb_5919 ACP was used solid-state NMR analysis as the kinetic data indicated that this form of the protein binds stronger to Bamb_5919 KS^0 (Figure 3.9). The samples that were prepared by mixing ^{15N}*holo*-Bamb_5919 ACP and Bamb_5919 KS^0 1:1 and 1:2 ratio (in the excess ^{15N}*holo*-Bamb_5919 ACP) did not yield a good enough NMR signal. This indicated that the sediment contained mostly a larger binding partner, Bamb_5919 KS^0. To push the equilibrium towards the complex formation, the ratio was increased to 5:1 which resulted in a detectable signal. In addition, sample pH was increased to 7.5 to avoid unwanted precipitation during sedimentation which impairs

sample homogeneity. An example of a proton spectrum recorded on such a sample is shown in Figure 3.36.

The sedimentation at 700000 g produced very solid gel sediment that was not feasible for manual packing into the rotors. The sedimentation at a lower speed, 500000 g, yielded a more viscous sample, which was also properly hydrated. Figure 3.36 shows the spectrum which was sedimented at that speed. In addition, to speed up the acquisition of experiments and thus enhance the signal, the samples were exposed to paramagnetic tag gadolinium diethylenetriaminepentaacetic acid bismethylamide (Gd(DTPA-BMA)). Namely, paramagnetic relaxation enhancing agents decrease ^{1}H T_1 relaxation time and allow for faster repetition of experiments and an increase in signal to noise ratio and are often used to speed up acquisition in solid-state NMR[146]. Gd(DTPA-BMA) is widely used in Lewandowski lab for protein sample preparation[154].

Ideally, the recorded CP spectrum would resemble an HSQC, i.e. all backbone residues involved in the formation of the Bamb_5919 ACP_Bamb_5919 KS^0 complex should give a signal. However, this has not yet been achieved for inhomogenous protein complexes such as ACP-KS^0. The 2D ^{1}H-^{15}N CP spectrum recorded at 60 kHz and 500 MHz showed the presence of several peaks that were possible to distinguish (Figure 3.37). The signals in ^{1}H-^{15}N CP originate only from immobilised residues within the protein backbone, i.e. those in the folded regions. Thus, only the residues in structured parts of the protein are likely to yield a peak in the spectrum [154]. Therefore, it was assumed that those signals originate from residues that are involved in the interaction with Bamb_5919 KS^0. Unfortunately, it was later discovered that the cooling unit on a spectrometer unexpectedly switched off during the experiment which heated up the sample to 50°C. Based on CD data (Figures 3.4 and 3.5), the increase in temperature probably did not have a profound effect on the stability of Bamb_5919 ACP. However, it most likely had a detrimental effect on the structure of Bamb_5919 KS^0.

.

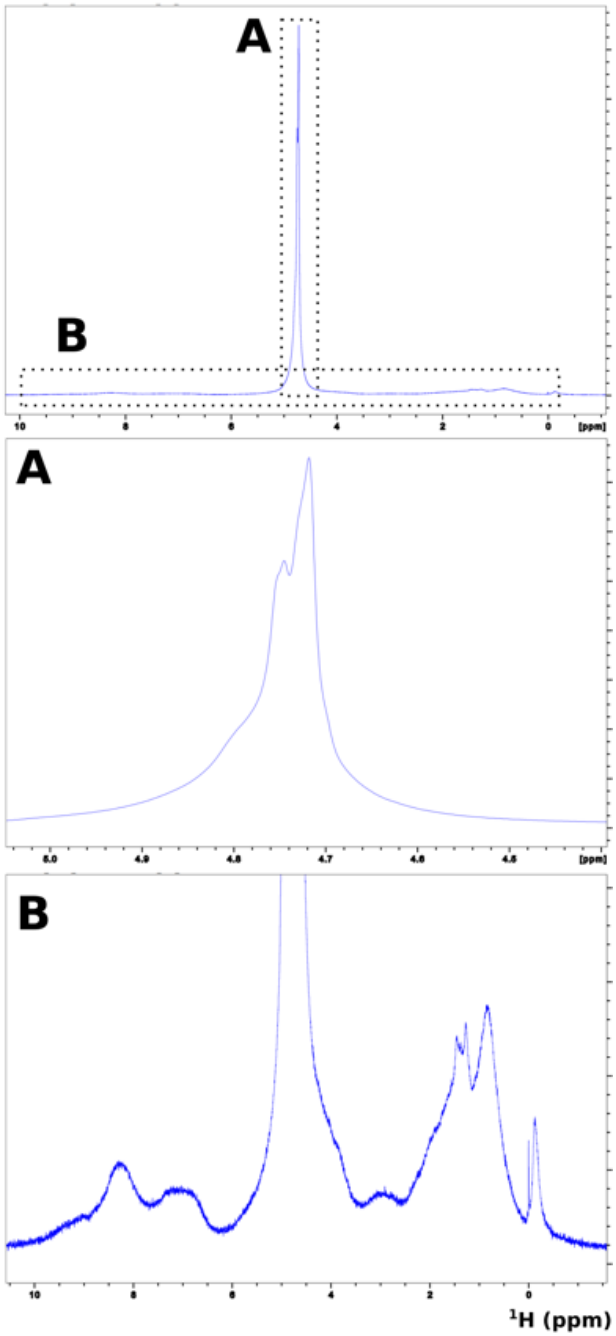


Figure 3.36 ^{1}H NMR spectrum of Bamb_5919 ACP in the complex with Bamb_5919 KS^0. The sample was well hydrated, as indicated from the water peak shown in A). The spectrum was recorded at 700 MHz. The sample temperature was 30^{0}C.

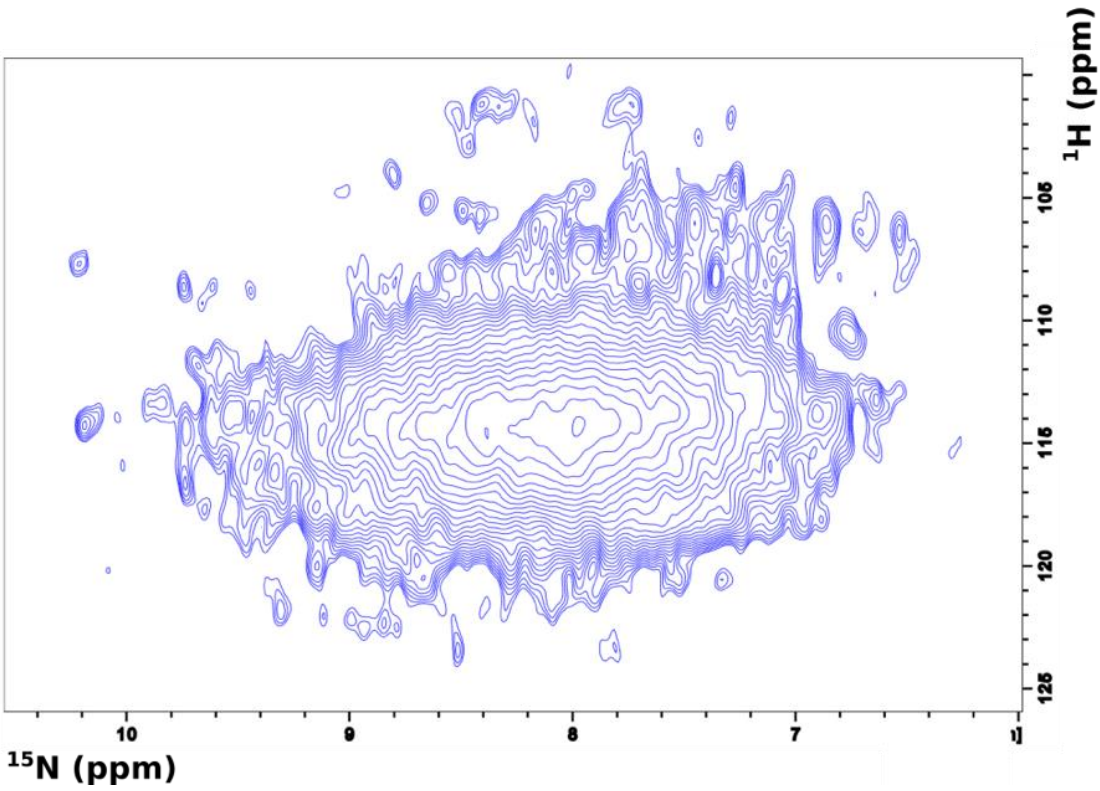


Figure 3.37 2D ^{1}H-^{15}N CP correlation spectrum of holo-Bamb_5919 ACP domain in a sedimented complex with Bamb_5919 KS^{0}. The sample was exposed to high temperature, which might have a negative effect on the protein fold of Bamb_5919 KS^{0}.

To improve spectral resolution, the sample was further analysed at higher MAS. The 2D ^{1}H-^{15}N CP spectra recorded at 90 kHz and 600 MHz also indicated that some peaks were possible to resolve (Figure 3.38). However, recording the same spectra with a higher number of scans did not yield the peaks with the same shift so it was not possible to reproduce the spectra and ensure that they are real peaks (Figure 3.39 – bottom spectrum). The overlaid 2D ^{1}H-^{15}N HMQC spectrum recorded for isolated *holo*-Bam_5919 ACP and ^{1}H-^{15}N CP spectra recorded for its complex with Bamb_5919 KS^{0} (Figure 3.38) show that CP spectrum covers the parts of HMQC spectrum that do not have any peaks. This implied that some peaks shifted significantly when in complex with Bamb_5919. This suggested that these peaks undergo a conformational change. This was in agreement with solution-state NMR data.

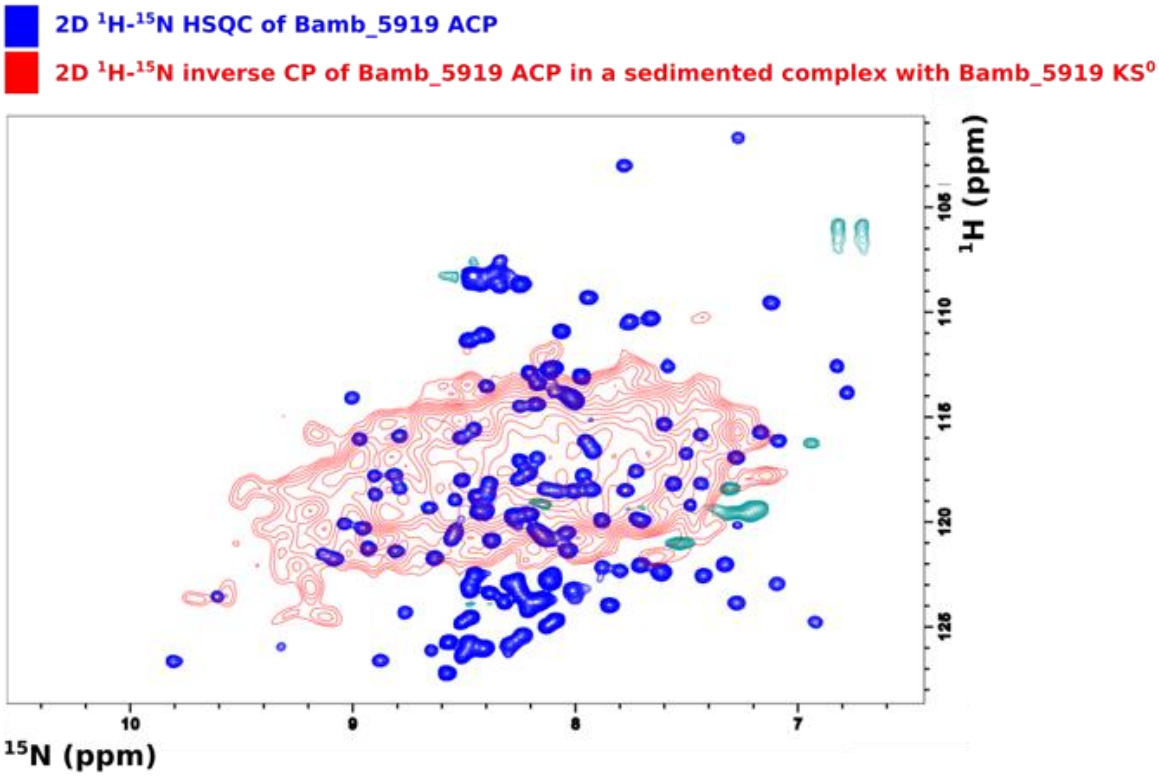


Figure 3.38 An overlay between 2D ^{1}H-^{15}N CP of holo-Bamb_5919 ACP domain in a sedimented complex with Bamb_5919 KS^{0} and unbound 2D ^{1}H-^{15}N HSQC. Spectrum was recorded at 600 MHz, 100 kHz.

Unlike CP, solution-state NMR-like experiment INEPT detects only signals originating from highly dynamical and flexible residues, which are usually located in terminal parts of the protein[154] that are most likely not rigidified by the presence of a binding partner. The INEPT spectrum of Bamb_5919 ACP:Bamb_5919 KS^{0} was overlapped with the HMQC spectrum of unbound ACP (Figure 3.39, spectrum in green). The INEPT peaks could be overlapped with the peaks in HMQC that were located in the flexible terminal parts of unbound Bamb_5919 ACP. This implied that they are not affected by complex formation, which was again in line with the results of solution-state NMR titrations. However, only one sedimented sample generated signals in the INEPT experiment so it was not possible to further elaborate on these conclusions and repeat the results.

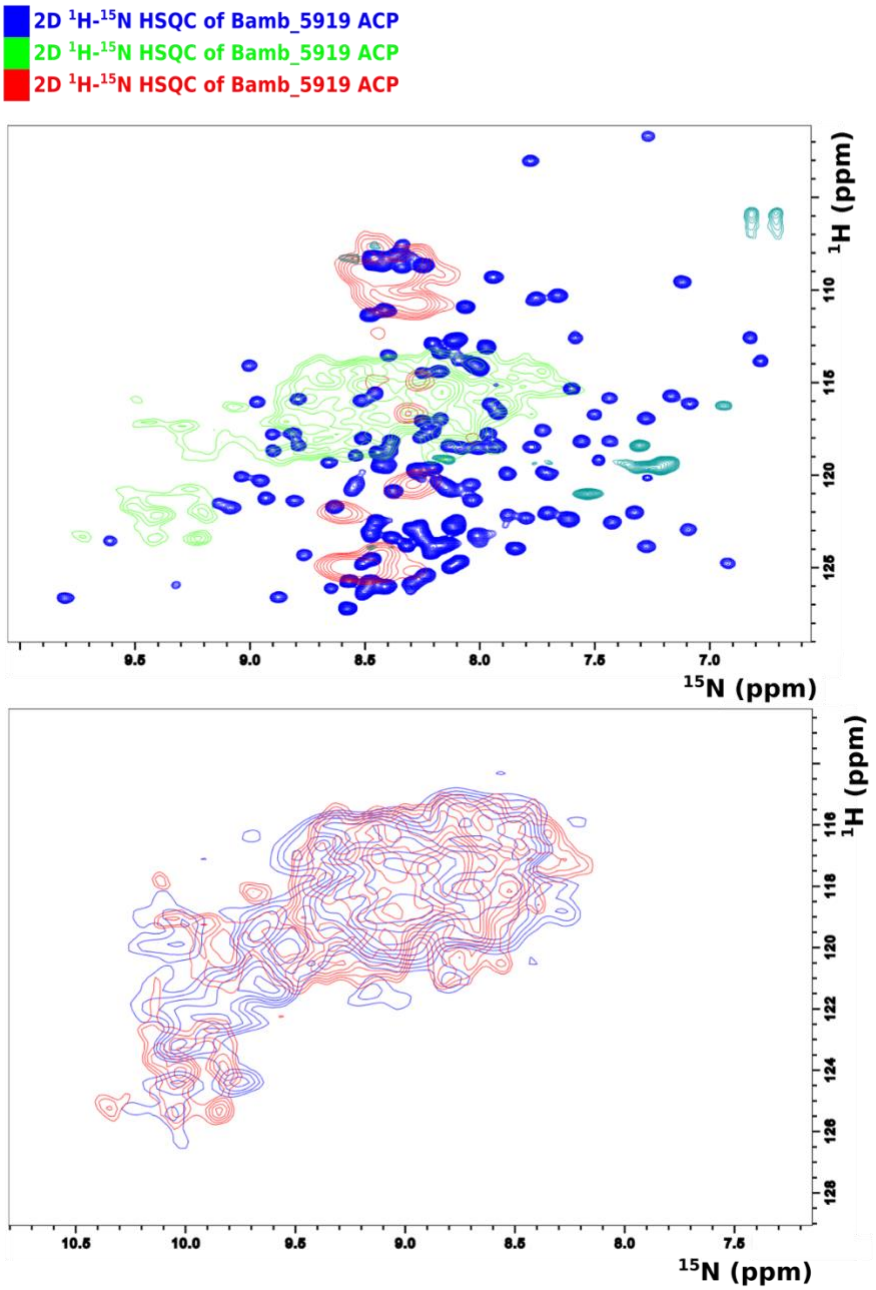


Figure 3.39 Figure 3.24 A) An overlay of 2D ^{1}H-^{15}N CP and INEPT of holo-Bamb_5919 ACP domain in a complex with Bamb_5919 KS^{0}. These solid-state NMR spectra were also overlapped with a solution-state ^{1}H-^{15}N HMQC of an isolated Bamb_5919 ACP. INEPT overlaps with the residues in HMQC spectrum which are in the disordered part of Bamb_5919 ACP (termini) B) Overlap between the two CP recorded with a time difference of several hours. The spectra cover the same region in CP, but the shifts of the distinguishable peaks are not identical, which means that it was not possible to reproduce the spectra and the individual peaks are not real. The solid-state spectra shown in the upper figure were recorded at 600 MHz, 90 kHz. The bottom CP spectra were recorded at 700 MHz, 100 kHz.

3.4 Conclusions

Majority of the assigned Bamb_5919 ACP backbone residues were predicted by ^{15}N relaxation experiments and TALOS+ analysis to be located in the parts of the protein that were categorised as helical by Phyre+ homology model. Solution-state NMR titrations indicated that only the residues located in helices are affected by the binding with Bamb_5919 KS^0, not N and C terminal parts of Bamb_5919 ACP. These residues were located in the core of the protein, which suggested that Bamb_5919 ACP undergoes conformational exchange upon binding to Bamb_5919 KS^0. Solid-state NMR analysis of the Bamb_5919 ACP:Bamb_5919 KS^0 interface further suggests that Bamb_5919 ACP changes its conformation in the complex, compared to its unbound form. Kinetic studies of the interface reveal that the interaction between the two interacting partners is strong.

Chapter 4 - The interaction of enacyloxin module 10 non-elongating ketosynthase and a downstream peptidyl carrier protein

4.1 Introduction

Chapter 3 provided information about the interface between Bamb_5919 ACP and Bamb_5919 KS^0. However, as a second step towards the understanding of what structural characteristics of enacyloxin module 10 ACP make it an improper interaction partner with the C domain, Chapter 4 provides an insight into the interaction between PCP and KS^0.

As discussed in Chapter 2, PCPs are small (~ 10 kDa) proteins that are characterized by a highly conserved signature motif of GxxSL/I. The serine residue in this motif is the site for 4′-phosphopantetheinylation. PCP domain shares the canonical four-helix arrangement with acyl carrier proteins from PKS systems. This domain can adopt multiple conformations, which is thought to stabilise specific conformations for a functional interaction[155].

The solution-state NMR structure of Bamb_5919 PCP (10.7 kDa) also reveals four α-helical fold that is typical for PCPs, with disordered regions of residues at both N- and C-terminus (Figure 4.1) [147]. To get insights into the interaction between Bamb_5919 PCP and Bamb_5919 KS^0, the approach used to analyse the interaction between Bamb_5919 ACP and Bamb_5919 KS^0 was employed and is presented in this chapter.

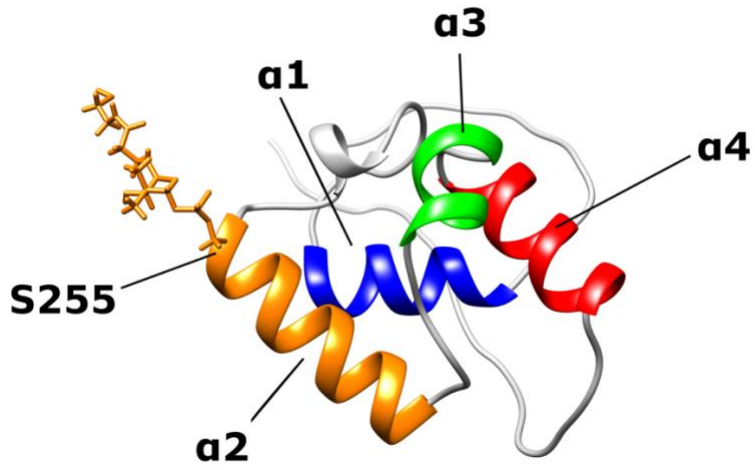


Figure 4.1 The NMR structure of holo Bamb_5917 PCP (PDB:5MTI)[147].

4.2 The analysis of the interaction between Bamb_5917 PCP and Bamb_5919 KS^0

4.2.1 The BLI analysis of Bamb_5917 PCP:Bamb_5919 KS^0 interaction

For comparative purposes, the reaction buffer was the same as the buffer used for NMR titrations of Bamb_5919 ACP and Bamb_5919 KS^0. Bamb_5917 PCP was titrated with increasing concentrations of Bamb_5919 KS^0. The binding was evident from the increase in signal upon addition of analyte (Supplementary S.3). Fitting of the data revealed the K_d of 0.19 ± 0.02 µM (an average value of the duplicate measurements) for the binding of *apo*-Bamb_5917 PCP and 0.18 µM ± 0.02 µM for *holo*-Bamb_5917 PCP (an average value of the triplicate measurements, Figure 4.2). This was an indication that the addition of prosthetic arm does not significantly affect the strength of the binding between the two forms of Bamb_5917 PCP with Bamb_5919 KS^0.

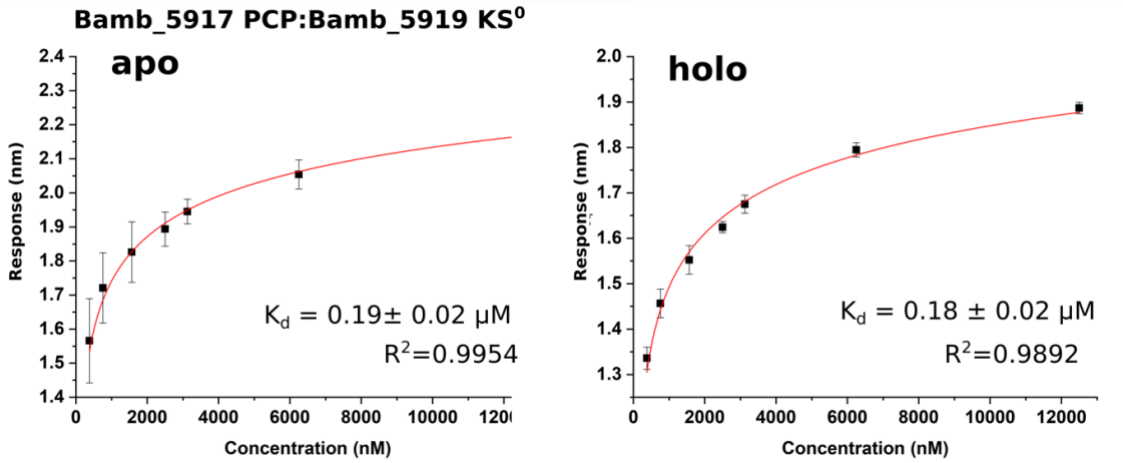


Figure 4.2. BLI steady-state analysis profile of the interaction between apo-Bamb_5917 ACP and holo-Bamb_5917 PCP with Bamb_5919 KS^0 (triplicate measurements). The binding curves can be found in Supplementary S.3.

The best data fitting was achieved with 2:1 fitting, which indicated non-ideal binding. The reasons that might underly this result are discussed in Chapter 3.

Next, ITC was used to get information about the binding enthalpy. The K_d value obtained from the fitting revealed a strong binding between Bamb_5919 PCP and Bamb_5919 KS^0, which was consistent with BLI data (Figure 4.2). The binding was also an endothermic reaction, which suggests that the binding is mainly mediated by hydrophobic interactions. However, the stoichiometry of binding was far off the 1:1 binding profile and there was a significant signal offset remaining after deduction of buffer signal.

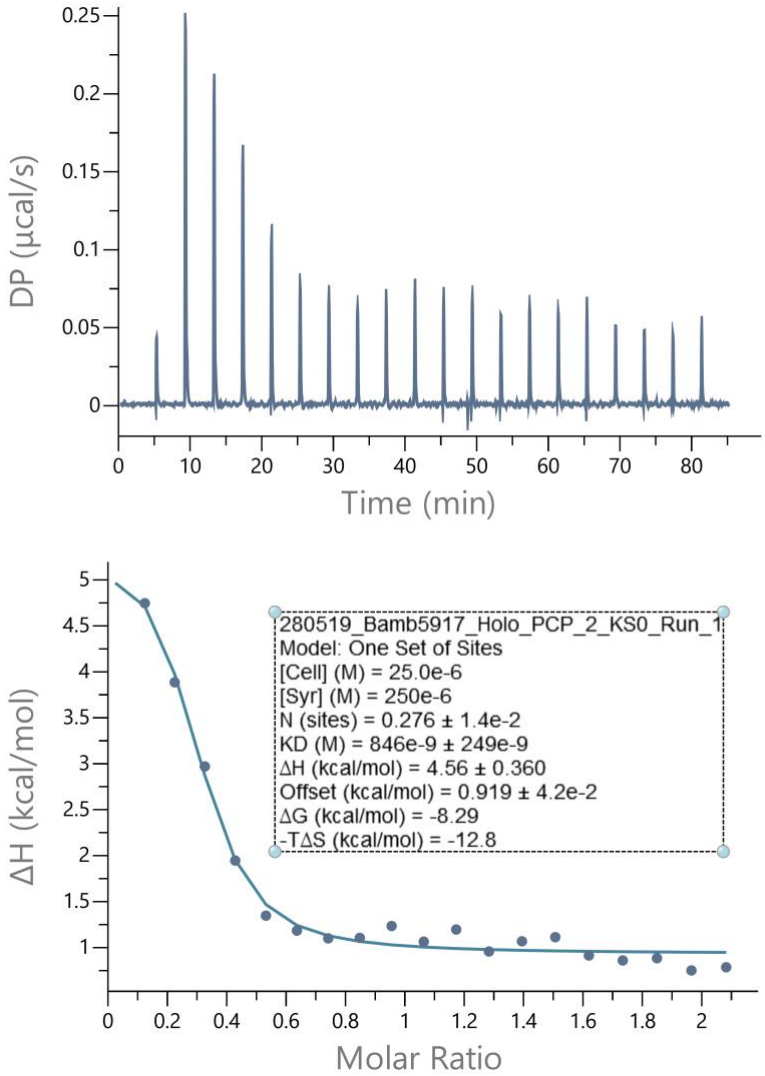


Figure 4.3 ITC analysis of the binding between Bamb_5919 ACP and Bamb_5919 KS^0. A) The measurement obtained for apo-Bamb_5919 ACP. B) The results obtained for two replicate apo-Bamb_5919 ACP measurements.

4.2.2 NMR analysis of apo to holo conversion of Bamb_5919 PCP

The difference between backbone chemical shifts in ^{1}H-^{15}N HMQC recorded for apo and holo-Bamb_5919 PCP was used to analyse the effect of phosphopantetheinilation on its backbone (Figure 4.4). The residues in the vicinity of helix 2 conserved serine, which acts as an acceptor site for Ppant, together with several residues in helix 3, were the only residues that were prominently shifted. This was most likely due to the local changes in electron density caused by the covalently bound phosphate and the pantetheine. The same effect was observed in NMR studies of aryl-acid-loaded PCP from yersiniabactin

NRPS[156] and the pyrrole-loaded PCP from pyoluteorin NRPS[156] which suggested that Ppant arm interacts with helices 2 and 3.

Besides, NMR structures of type I PCPs isolated from the NRPSs revealed three interchanging conformations, the A state for apo-PCP (without Ppant), H state for holo-PCP and a conformation shared by both apo- and holo-PCP, the A/H state[155,157].

4.2.3 Investigating the interaction between Bamb_5917 PCP and Bamb_5919 KS^0 by solution-state NMR titrations

To get insights into the effect of the binding with Bamb_5919 KS^0, a series of ^{1}H-^{15}N-HMQCs of apo- and holo-^{15}N-Bamb_5917 PCP were recorded in the presence of increasing concentrations of unlabelled Bamb_5919 KS^0. The spectra of unbound and bound Bamb_5917 PCP are shown overlaid in Figure 4.5. In both spectra, some residues were in a slow exchange upon addition of Bamb_5919 KS^0. These peaks were tracked and their changes quantified as a function of K_d (Figure 4.6) to map the binding interface. There were no overlapping peaks so it was possible to extract integrals for all peaks. The absence of newly formed signals suggested that the complex formation resulted in line broadening as the size of ^{15}N-Bamb_5919 PCP:Bamb_5919 PCP complex was 85 kDa, compared to 10.7 kDa of the unbound protein. The extracted local K_d values were consistent across replicates for both apo and holo samples (Figure 3.6).

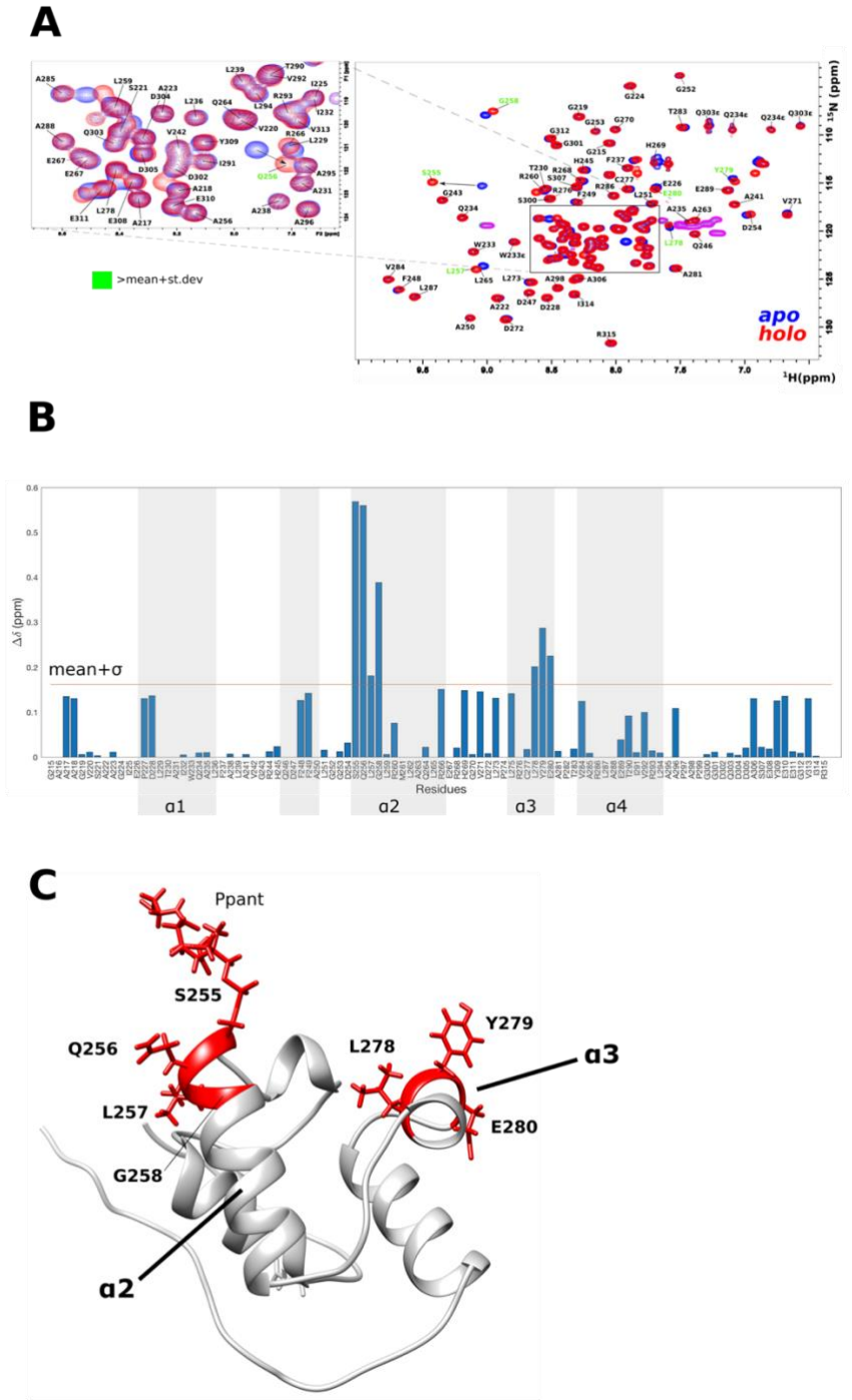


Figure 4.4 Chemical shift difference between apo and holo form of Bamb_5917 PCP. A) The overlap between ^{15}N-HSQC spectra recorded for apo and holo forms. B) Average chemical shift differences, $\Delta\delta_{av}$, shown between holo and the apo form, C) The residues with $\Delta\delta_{av}$ above the average are shown on PCP model.

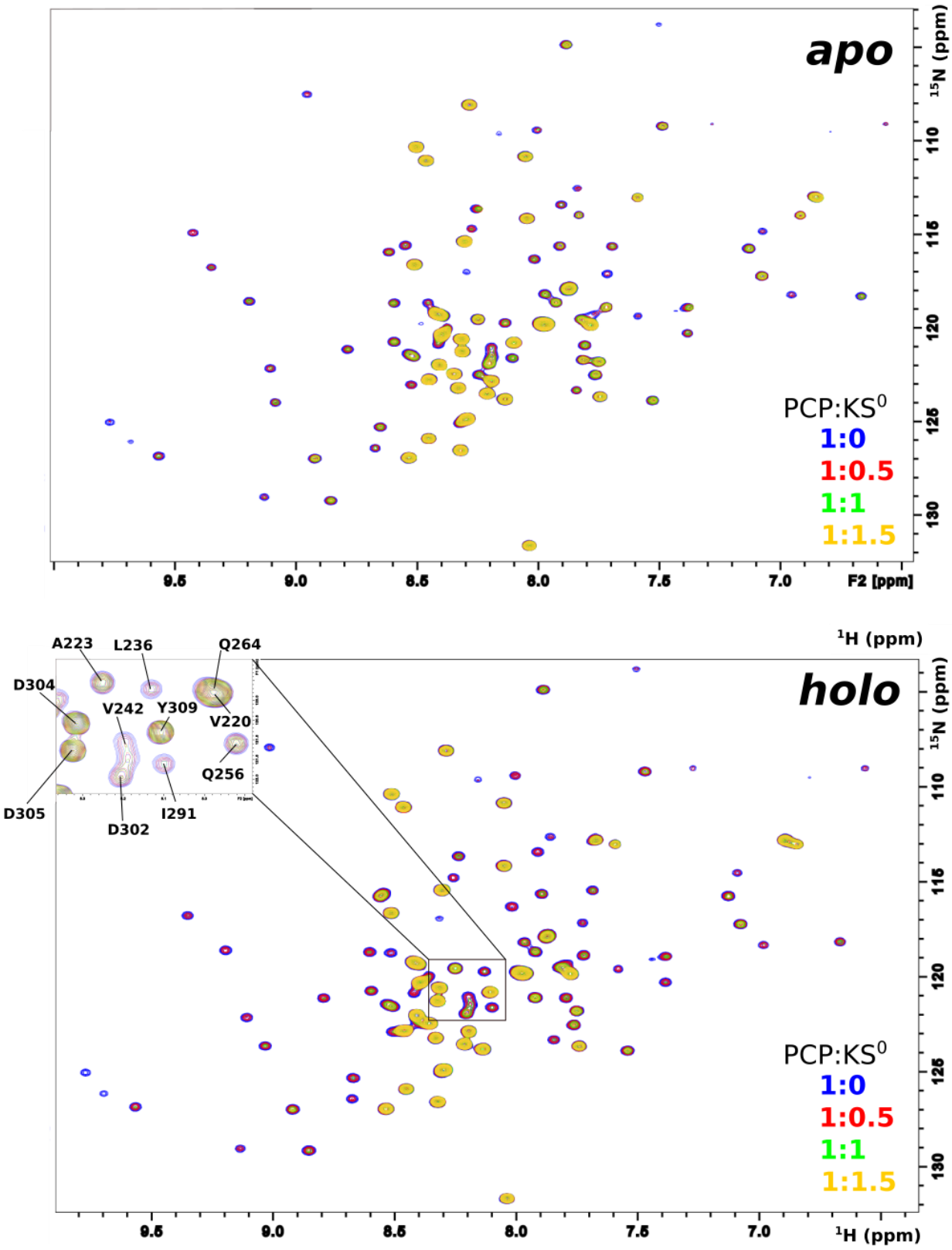


Figure 4.5 ^{15}N-^{1}H HMQC spectra of titration of apo and holo-Bamb_5919 PCP with increasing concentrations of Bamb_5919 KS^0.

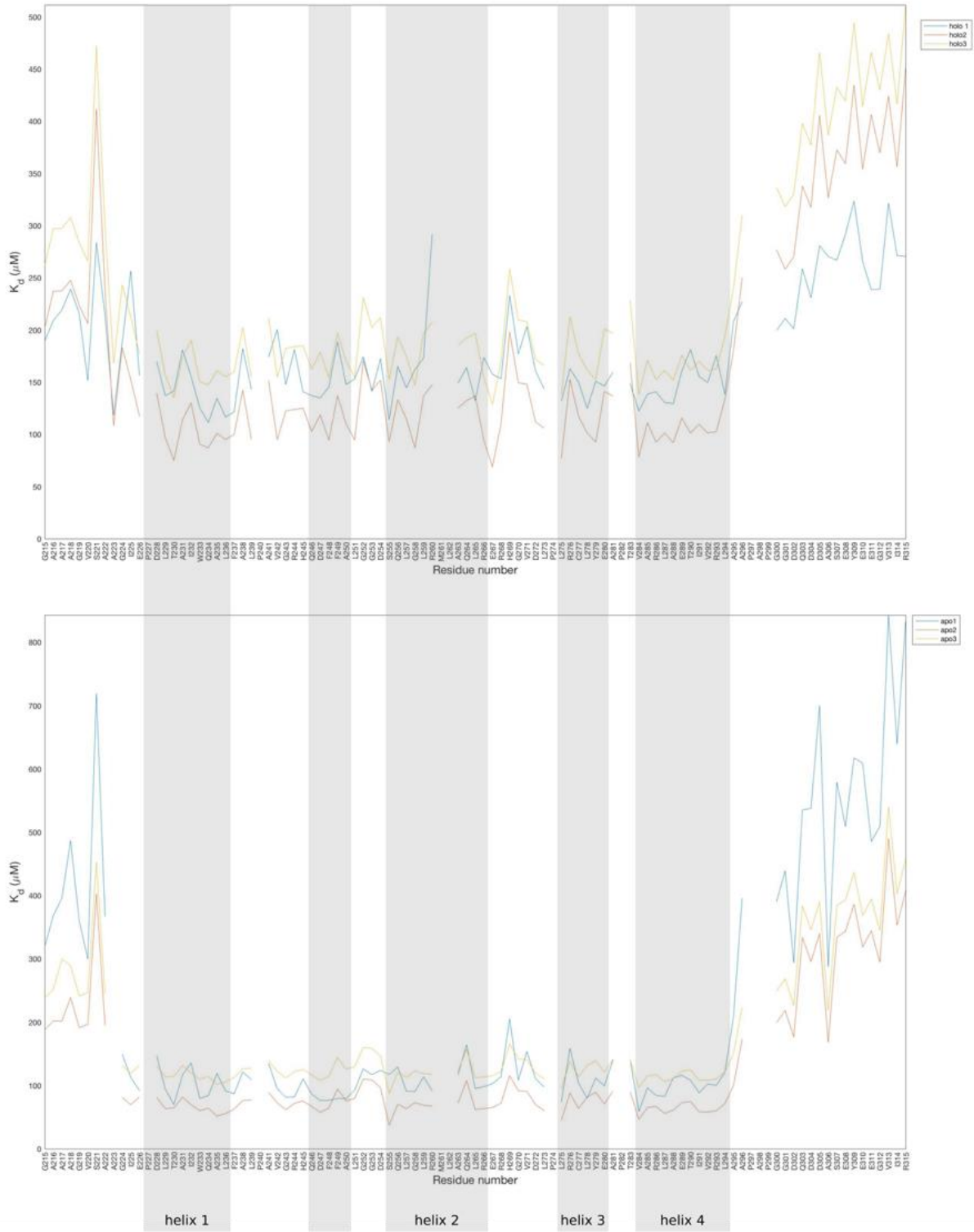


Figure 4.6 Local apparent association constants, K_d, obtained from titrations of apo and holo-Bamb_5917 PCP with Bamb_5919 KS^0. The gaps represent unassigned residues.

The apo-Bamb_5917 PCP residues that were significantly affected by the binding to Bamb_5919 KS^0 were located in helices (Figure 4.7). However, apart from being oriented towards the protein core, they seemed to form a surface patch. In order to inspect whether the surface contacts might be involved in the binding, the data recorded for holo-Bamb_5919 PCP were also analysed (Figure 4.8). This analysis, however, showed that the residues with the highest peak intensity decrease are located in the core of the protein and do not form a surface patch. This indicated that holo-Bamb_5917 PCP undergoes a conformational change upon binding with Bamb_5919 KS^0, which is consistent with a highly dynamic nature of PCPs who are known to undergo conformational rearrangements when interacting with their domain partners[158,159].

For example, the study of PCP-TE didomain protein from *E. coli* enterobactin synthetase EntF NRPS subunit[160] and a protein complex between a type II TE, SrfTEII, from *Bacillus subtilis* NRPS and Type I PCP from the third module of the tyrocidine A synthetase[161] also led to the conclusion that the PCP and/or the catalytic domains undergo large conformational changes upon interaction.

It is not possible to conclude from the presented results whether the contact between apo-Bamb_5917 PCP is guided by the surface patch in any way. It clear, however, that the unstructured parts of Bamb_5917 PCP probably do not interact with KS^0.

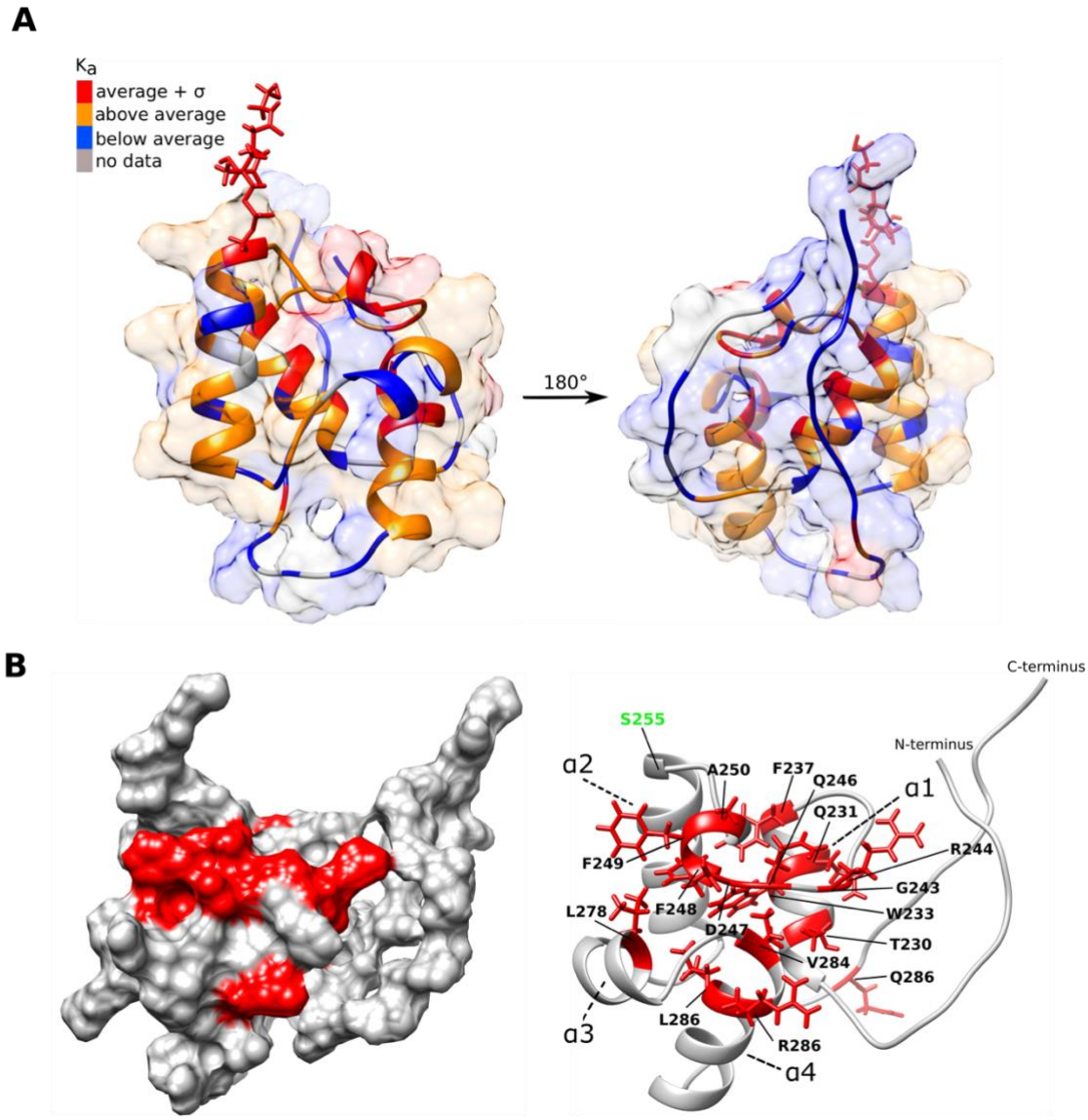


Figure 4.7 Local apparent Ka values projected onto the structure of apo-Bamb_5919 PCP. The residues highlighted in red have a K_a that is higher than one standard deviation based on the obtained data from the analysis of the titrations of apo-Bamb_5917 PCP with Bamb_5919. KS^0.

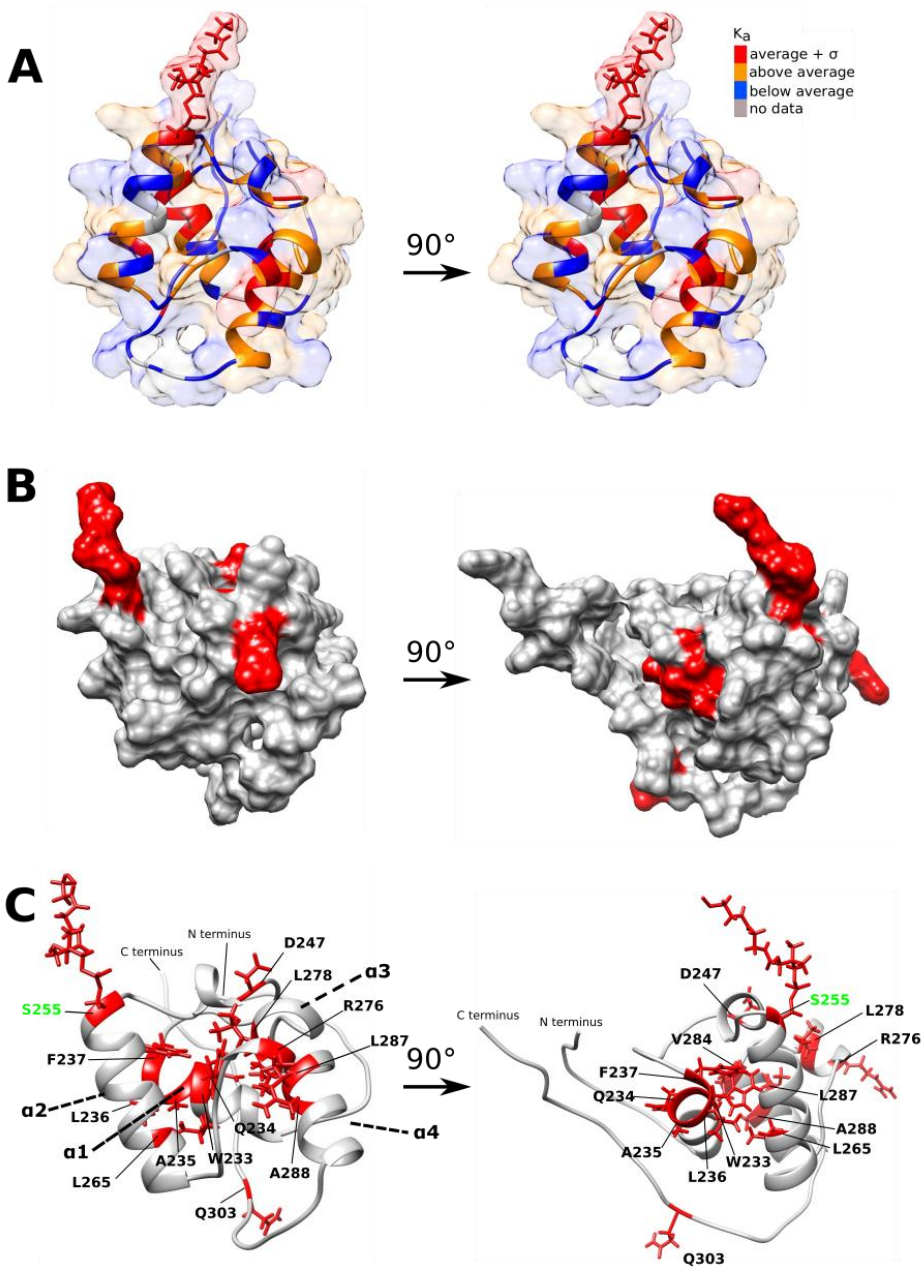


Figure 4.8 Local apparent Ka values projected onto the structure of holo-Bamb_5919 PCP. The residues highlighted in red have a K_a that is higher than one standard deviation based on the obtained data from the analysis of the titrations of holo-Bamb_5917 PCP with Bamb_5919. KS[0]

4.3 Conclusions

Based on solution-state NMR titrations, the interaction between Bamb_5917 PCP and Bamb_5919 KS^0 follows a slow NMR exchange regime, which is a characteristic of strong binding. This was also confirmed by BLI and ITC experiments. The residues that are most affected by the interaction are located in the protein core, which indicates conformational rearrangement upon binding.

Chapter 5 - Analysis of the structural differences between Bamb_5919 ACP and Bamb_5917 PCP which affect their interaction with the C domain

5.1 Introduction

Chapter 3 and 4 focused on the interaction profiles of enacyloxin module 10 ACP:KS^0 and KS^0:PCP separately. This chapter compares the results obtained from both of the chapters and suggests which structural characteristics of ACP make it an inappropriate binding partner with C domain, also based on a recently published Lewandowski/Challis lab analysis of the PCP-C domain interface[147]. It has been mentioned in Chapter 3 that the structure of module 10 KS^0 has not yet been determined, so this chapter will only discuss the binding interface from a perspective of carrier proteins.

Besides, competitive NMR titrations were used to determine whether PCP and ACP bind KS^0 simultaneously or competitively, i.e. at the same or different location.

5.2 Competitive NMR Titrations

Firstly, unlabelled Bamb_5919 KS^0 was titrated into the solution of *holo*-^{15}N-Bamb_5919 ACP and the backbone NH groups monitored using HMQCs (Figure 5.1). Unlabelled *holo*-Bamb_5917 PCP was then added to the complex but no changes were seen, suggesting that Bamb_5919 ACP did not bind Bamb_5919 KS^0 in the presence of Bamb_5919 PCP (Figure 5.1). Firstly, the intensity of some of the peaks decreased upon binding to Bamb_5919 KS^0. This was expected as it was, determined in Chapter 3 that Bamb_5919 ACP interacted with Bamb_5919 KS^0, which resulted in decreased intensity of some of the peaks. However, those peaks did not regain its intensity even after addition of a 3-fold excess of Bamb_5919 PCP, indicating that the Bamb_5919 ACP was not displaced by Bamb_5919 PCP on Bamb_5919 KS^0 interface. These experiments were recorded at sufficiently high protein concentrations (100 µM ^{15}N-Bamb_5919 ACP, 100 µM Bamb_5919 KS^0, 200 µM Bamb_5919 PCP) to be more than the K_ds of the individual interactions (Bamb_5919 ACP:Bamb_5919 KS^0 K_d = 0.23 µM, Bamb = 0.18 µM). The reverse titrations further confirmed that Bamb_5919 ACP and Bamb_5919 PCP did not bind competitively (Figure 5.2).

Visual inspection of HMQC spectra shown in Figure 5.1 and 5.2 was followed up by the analysis of residue peak intensities throughout titrations experiments (Figures 5.3 and 5.4).

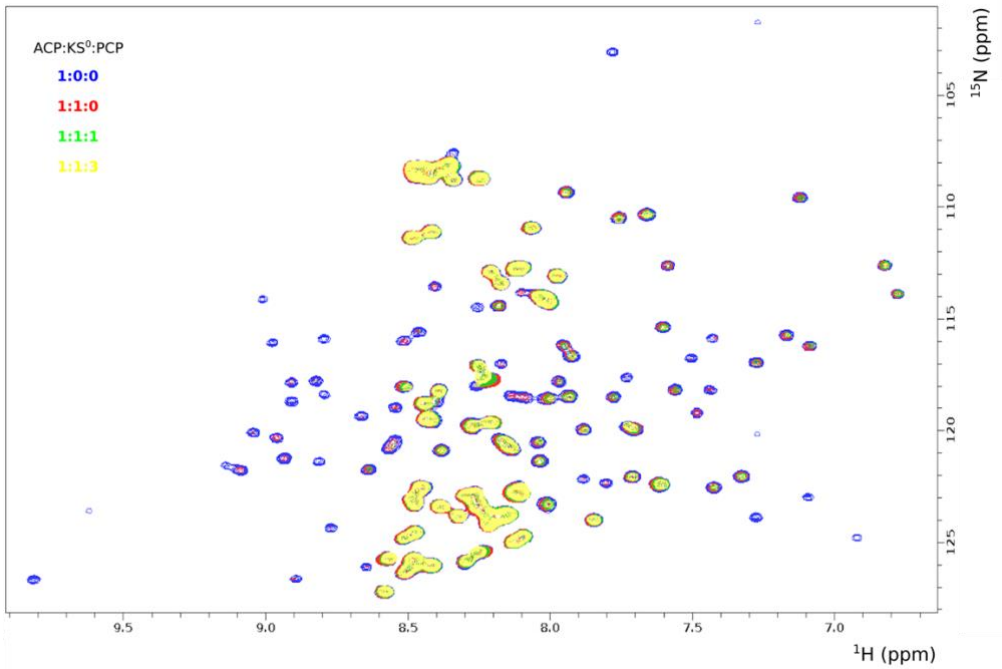


Figure 5.1 ^{1}H-^{15}N HMQC titrations of holo-Bamb_5919 ACP with the increasing concentrations of Bamb_5919 KS^{0}. The addition of ligand caused some peaks to decrease in intensity. The intensity of these peaks did not increase upon addition of unlabelled Bamb_5917 PCP, which indicated that Bamb_5919 ACP does not bind in the same location on Bamb_5919 KS^{0}.

The analysis showed that the addition of Bamb_5919 KS^{0} caused intensity decay of some of the peaks. However, upon addition of an unlabelled carrier protein there was no change in the peak intensity. Again, this showed that Bamb_5919 ACP and Bamb_5917 PCP most likely do not share the interaction interface on Bamb_5919 KS^{0}.

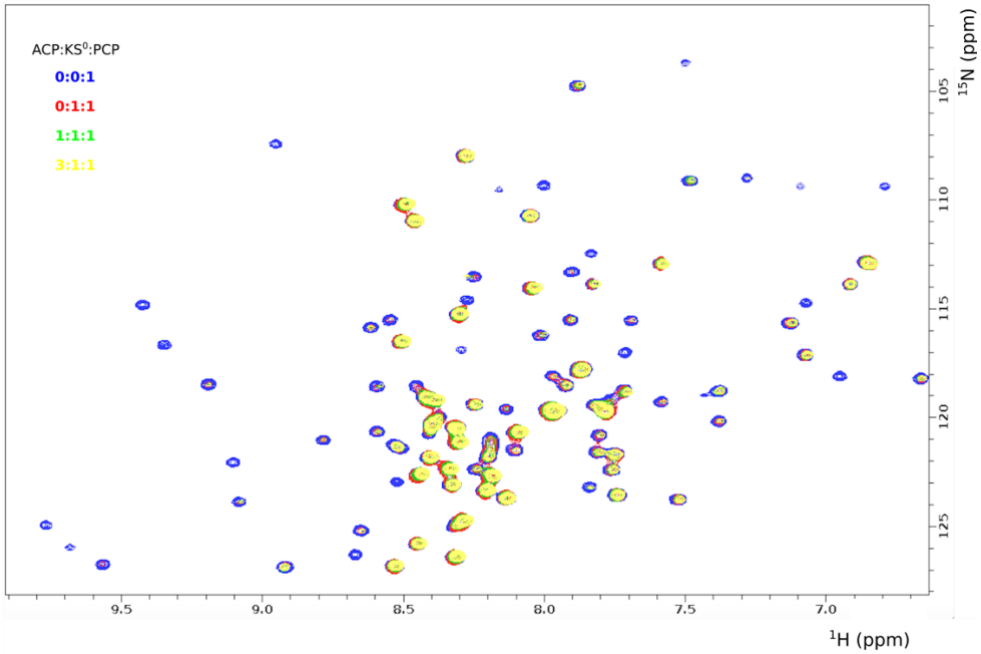
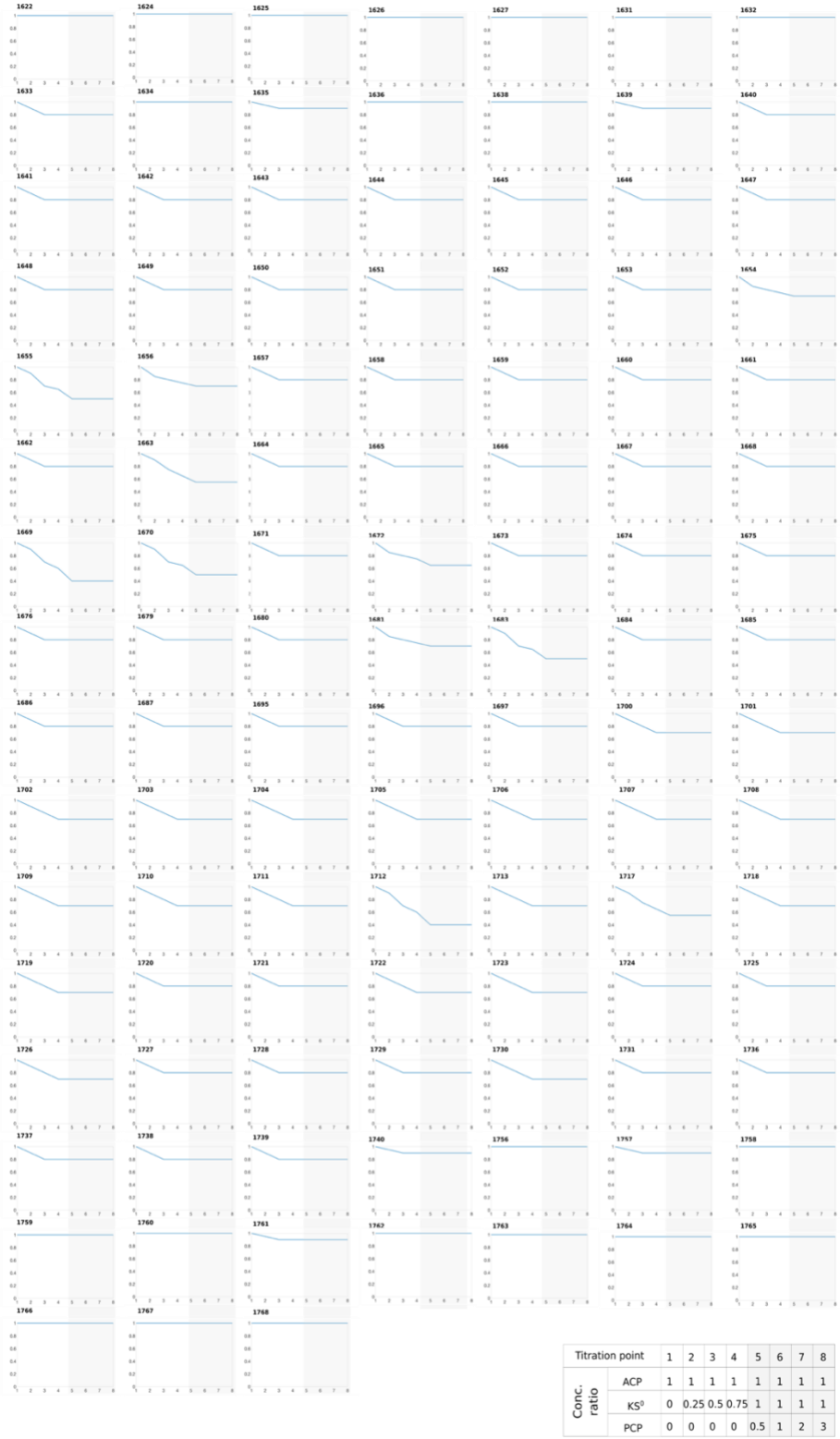
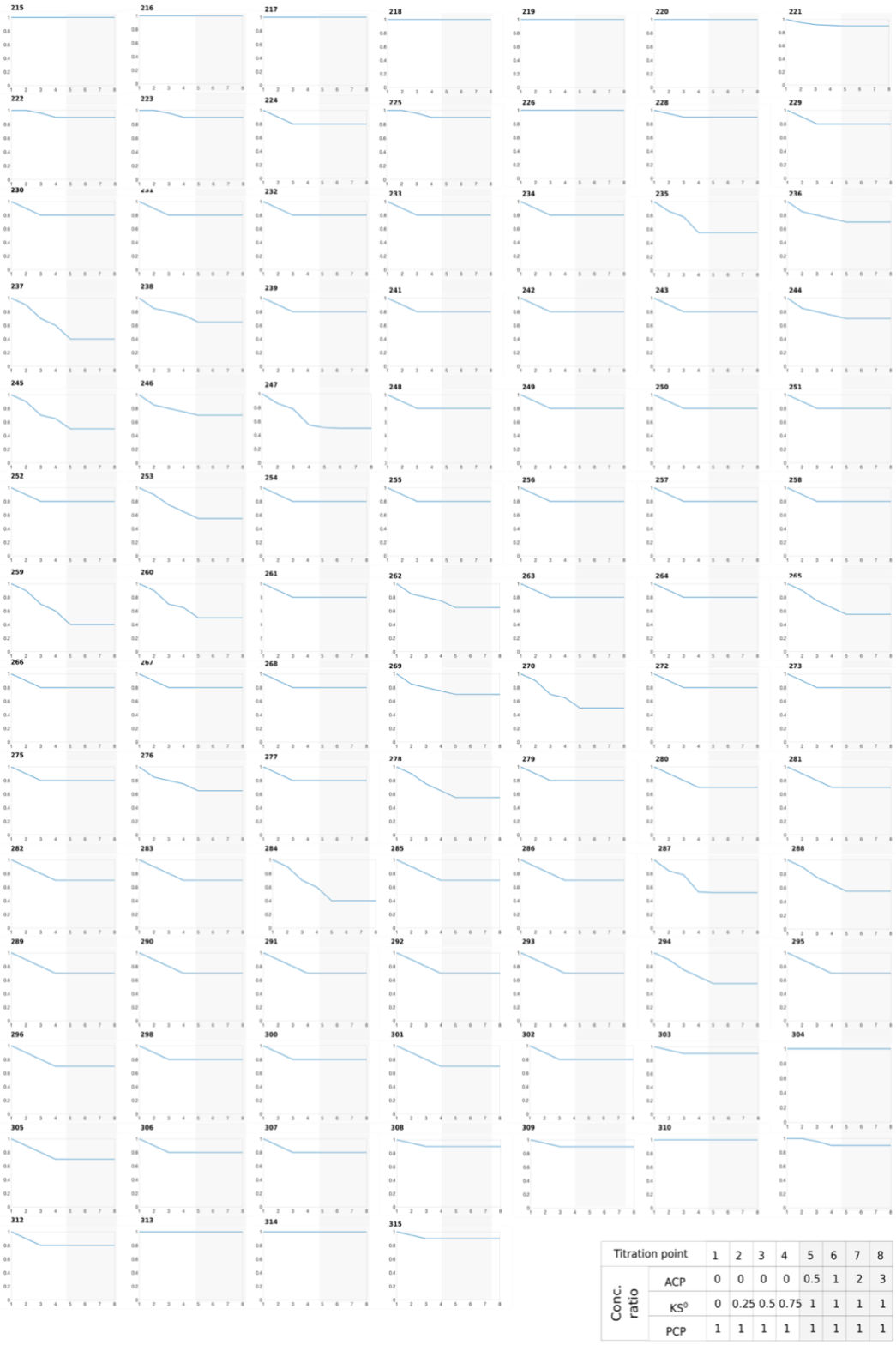


Figure 5.2 Figure 0.1 ^{1}H-^{15}N HMQC titrations of holo-Bamb_5917 PCP with the increasing concentrations of Bamb_5919 KS^{0}. The addition of ketosynthase caused the intensities of some of the peaks to decrease. However, they were not recovered upon addition of unlabelled Bamb_5919 ACP, which confirmed that Bamb_5919 ACP and Bamb_5917 PCP do not bind Bamb_5919 KS^{0} at the same location.



Titration point		1	2	3	4	5	6	7	8
Conc. ratio	ACP	1	1	1	1	1	1	1	1
	KS^0	0	0.25	0.5	0.75	1	1	1	1
	PCP	0	0	0	0	0.5	1	2	3

Figure 5.3 NMR Titrations of holo-^{15}N-Bamb_5919 ACP with Bamb_5919 KS^0 and Bamb_5917 PCP. The intensity decay rates are shown for individual residues. The dilution effect was taken into account. The increase in Bamb_5919 PCP concentration (highlighted in gray) does not affect the NMR signal decay, indicating that PCP and ACP do not bind in the same site of KS^0.



Titration point		1	2	3	4	5	6	7	8
Conc. ratio	ACP	0	0	0	0	0.5	1	2	3
	KS^0	0	0.25	0.5	0.75	1	1	1	1
	PCP	1	1	1	1	1	1	1	1

Figure 5.4 NMR Titrations of holo-^{15}N-Bamb_5917 PCP with Bamb_5919 KS^0 and unlabelled Bamb_5919 ACP. The intensity decay rates are shown for individual residues and the dilution effect was considered in the calculations. The addition of Bamb_5919 ACP (highlighted in gray) does not have an effect on the decay rate of holo-^{15}N-Bamb_5917 PCP NMR signals, further suggesting that, together with the data presented in Figure 5.3, PCP and ACP do not bind in the same location on KS^0.

5.3 Comparison of binding interfaces formed by Bamb_5919 ACP, Bamb_5919 KS^0. Bamb_5917 PCP and Bamb_5915 C

The crystal structure of C domain had been determined by the collaborating prof. Tsai group[147].

This crystal structure was used for docking with the NMR structure of Bamb_5917 PCP[126]. The resulting docked complex was used in the chapter to compare the analysed binding interfaces.

Thus, homology model of apo-Bamb_5919 ACP (Figure 3.1) was overlaid to the NMR structure of holo-Bamb_5919 PCP, which was docked into the active site of the crystal structure of C domain (Figure 5.4). The superposition in Figure 5.4, together with the sequence alignments shown in Figure 5.5 and Figure 5.6, revealed that the major difference between the two carrier proteins is the length of helix 1. Helix 1 in Bamb_5919 ACP (20 residues based on the homology model in Figure 3.1) is longer than helix 1 of Bamb_5919 PCP (10 residues). Furthermore, the ^{15}N backbone relaxation data and TALOS+ analysis discussed in Chapter 3 suggest that helix 1 of Bamb_5919 ACP is even longer than predicted from the homology model. A long helix 1 might negatively interfere with the likelihood of Bamb_5919 ACP to properly fit into the active site of the C domain (Bamb_5915 C).

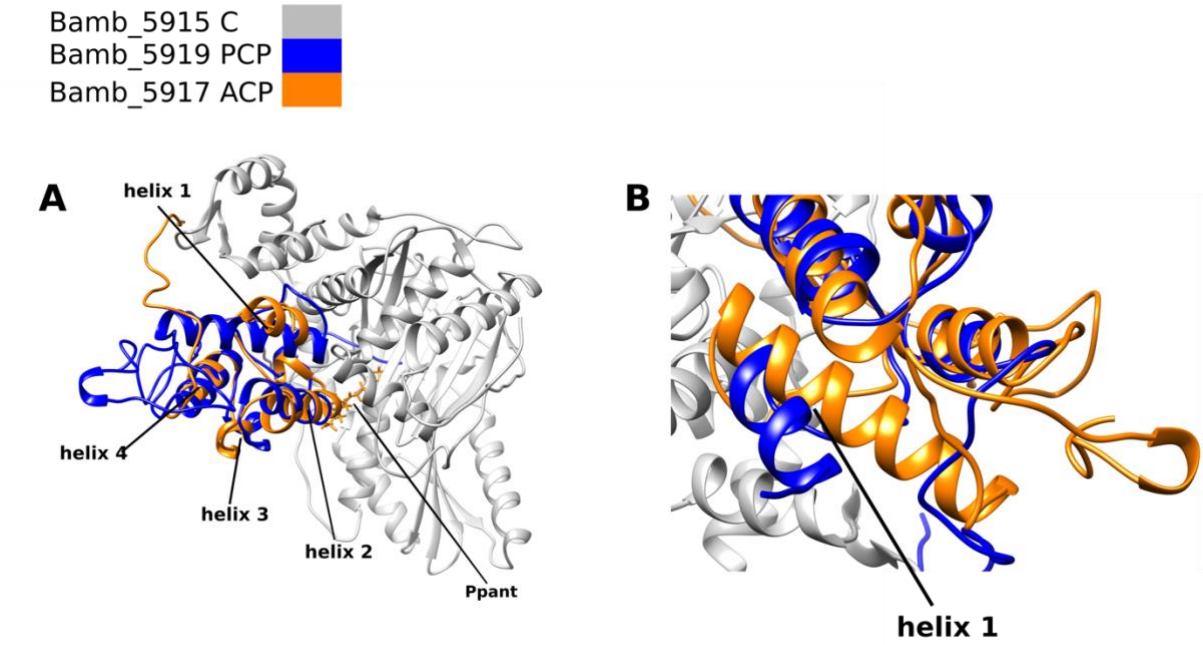


Figure 5.4 Bamb_5917 PCP docked into Bamb_5915 C139. Homology model of Bamb_5919
Homology model of Bamb_5919 ACP is superimposed with Bamb_5917 PCP. B. Helix 1 of Bamb_5919 ACP is longer than the same helix of Bamb_5917 PCP.

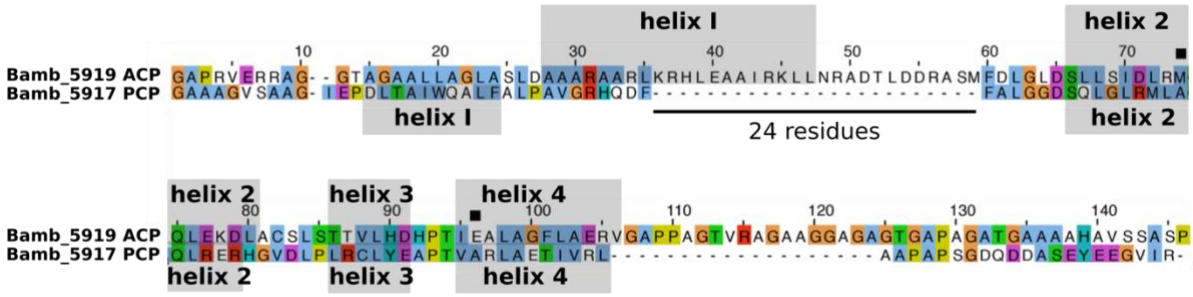


Figure 5.5 Sequence alignment between Bamb_5919 ACP and Bamb_5917 PCP. Sequence alignment was performed in Clustal Ω. Helix 1 of Bamb_5919 ACP is longer than the same helix in Bamb_5919 PCP.

Furthermore, the sequence alignment between the two carrier proteins reveals that several amino-acids, mostly in helices 2, and 4 differ significantly in charge (Figure 5.6, residues in rectangles). To inspect whether these residues affect the interaction of Bamb_5919 ACP with C domain, the Bamb_5919 ACP residues that were affected by the

presence of Bamb_5919 KS^0 were also shown in Figure 5.6 (red squares, taken from Figure 3.30). Majority of the Bamb_5919 ACP residues that were highly affected by the binding with Bamb_5919 KS^0 were not significantly different in terms of charge or hydrophobicity to the aligned residues from Bamb_5919 PCP. This indicated that these differences do not explain why Bamb_5919 ACP cannot bind to C domain.

Figure 5.6 Schematic representation of the sequence alignment between Bamb_5919 ACP and Bamb_5917 PCP. Red squares represent the residues that were affected during titrations of Bamb_5919 ACP with Bamb_5919 KS^0. Red rectangles pinpoint the residues that differ in charge between the two sequences.

However, the docking between Bamb_5919 PCP and Bamb_5915 C also showed that the C-terminal part of Bamb_5919 PCP aligns with the β-hairpin docking domain (βHDD) located at the N-terminus of C domain of Bamb5915 (Figure 5.8B). The C-terminal part of Bamb_5919 PCP is a glutamate-rich intrinsically disordered motif, which is also called SLiM (Short Linear Motif).

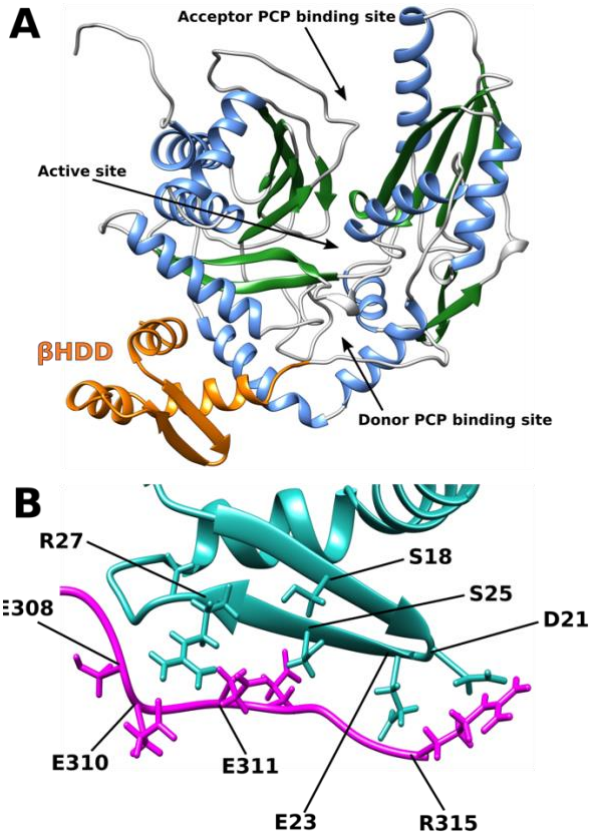


Figure 5.7 The structure of Bamb_5919 C (A). The beta-hairpin docking domain (βHDD) is shown in the figure and highlighted in orange. Bamb_5919 PCP is not shown here. (B) The position of SLiM relative to the docking domain with the indicated charged and polar residues that might mediate the interaction.

Such linear motifs (usually 3-10 amino acids long) are a common functional module within intrinsically disordered proteins. They regulate low-affinity interactions and can target proteins to a particular subcellular location, recruit enzymes that alter the chemical state of the motif by post-translational modifications (PTMs), control the stability of a protein, and promote the recruitment of binding factors to facilitate complex formation[162–165].

The docking domains, such as βHDD, are located at the terminal parts of the modules within PKSs and NRPSs and mediate the communication between the modules within PKSs and NRPSs[166,167]. More precisely, they are located at the C-terminus of an upstream ACP/PCP subunit and the N-terminus of a downstream KS or C subunit.

These domains are crucial for the interaction and it has been demonstrated that the use of genetic engineering to create mismatched docking domains significantly decreased the rate of biosynthesis[168,169].

The charged Bamb_5917 PCP SLiM residues probably establish contacts with the charged amino acids of βHDD (Figure 5.7), which facilitate the interaction between Bamb_5919 PCP and Bamb_5915 C.

Finally, a comparison of the NMR titrations results obtained for the interactions between Bamb_5919 PCP and Bamb_5919 KS^0 further suggest that the SLiM motif is required for the proper interaction between with Bamb_5915 C (Figure 5.8). Figure 5.8 shows that the profile of Bamb_5917 PCP residues being affected by the presence of Bamb_5919 KS^0 (Figure 5.8A) and Bamb_5915 C (Figure 5.8B) is similar in helical parts of the protein. The titrations of Bamb_5917 PCP and Bamb_5915 were obtained by Dr. S. Kosol. However, it is clear that SLiM motif is not affected by the presence of Bamb_5919 KS^0. Thus, the data suggest that charged residues in SLiM motif are important to establish contact with βHDD. The C-terminus of ACP is mostly hydrophobic, which might prevent it from interacting with the docking domain. It is rather difficult to speculate which type of interactions might mediate the interaction between Bamb_5919 KS^0 and both carrier proteins as the crystal structure of Bamb_5919 KS^0 is not complete (the alignment between Bamb_5919 KS^0 and ketosynthases from other systems is shown in Fig. S-4.5). However, the NMR data presented in Chapters 3 and 4 suggests that this might be driven by hydrophobic interactions being established between the carrier proteins and KS^0 upon binding.

5.4 Conclusions

Chapter 5 suggests that Bamb_5919 ACP and Bamb_5917 PCP do not bind in the same location on Bamb_5919 KS^0. The alignment between homology model of apo-Bamb_5919 ACP and holo-Bamb_5917 PCP docked with Bamb_5915 C indicates that the length of Bamb_5919 ACP helix 1 might prevent it from binding into the active site of Bamb_5915 C. Furthermore, NMR titrations discussed in Chapters 3 and 4 emphasize that the glutamate-rich SLiM motif might make Bamb_5917 PCP a preferential binding partner with Bamb_5915 C docking domain βHDD over Bamb_5919 ACP, which lacks charged residues in its C-terminus.

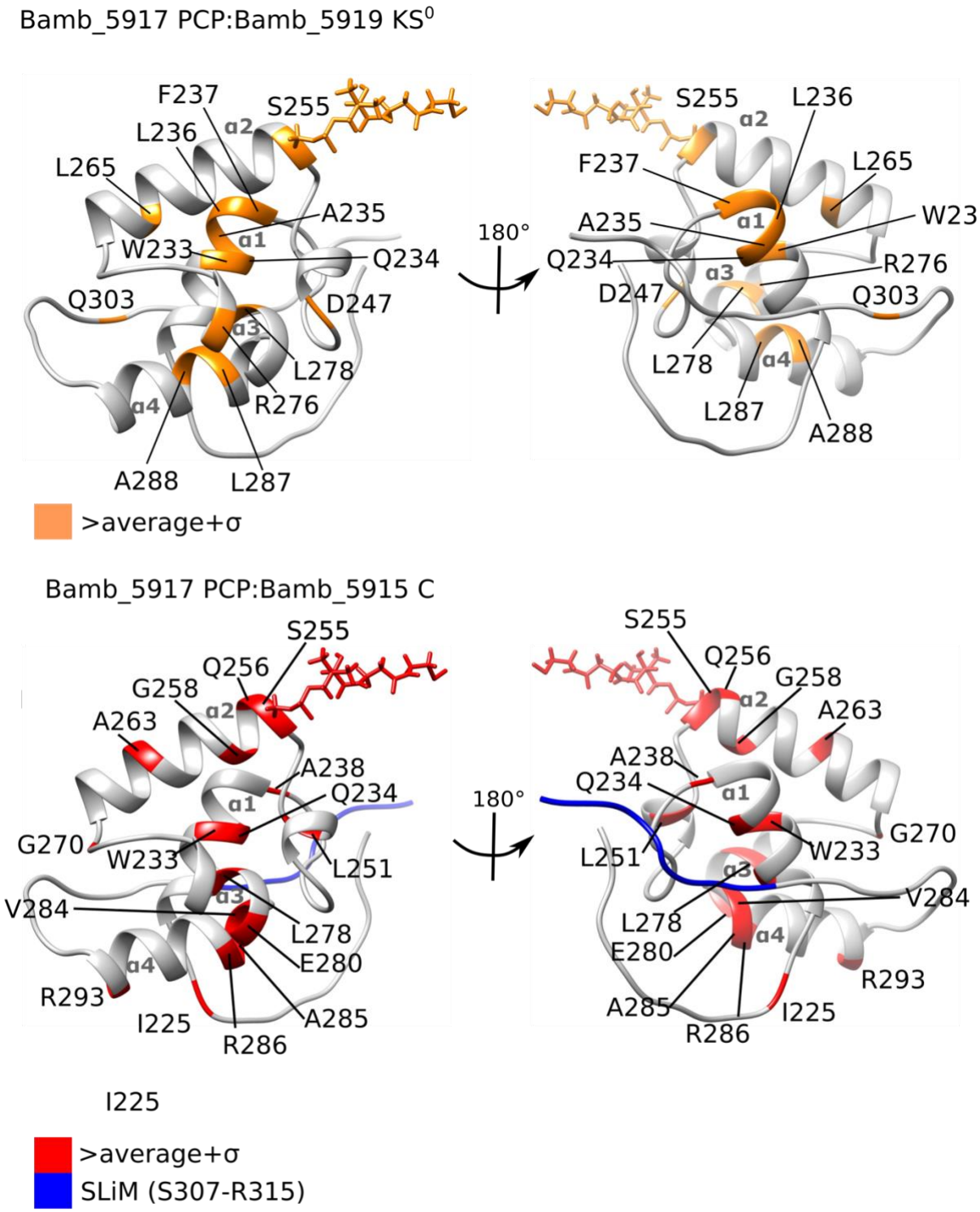


Figure 5.8 The Bamb_5917 PCP residues that were the most affected by titrations with Bamb_5919 KS0 (A) and Bamb_5915 C(B). The schematic representation of the alignment between the C-terminus of Bamb_5919 ACP and SLiM is also shown. Charged residues are highlighted in grey.

Conclusions and future work

From the perspective of bacterial cell, incorporation of non-elongating synthases into PKS and NRPS pathways might appear to be an energetically costly mechanism of overcoming domain-domain interaction incompatibilities. Namely, non-elongating ketosynthases are large proteins and their synthesis and degradation is an ATP consuming process. Thus, is peculiar that the evolution of biosynthetic pathways was directed towards incorporation of such enzymes, instead of mutations of domain-domain interface residues to establish favourable interactions. Nevertheless, it seems that non-elongating synthases are utilised by a vast number of PKS and NRPS systems to direct the biosynthesis of natural products. The work presented in this thesis attempted to understand what particular aspects of interaction between the enacyloxin IIa chain termination modules ACP, KS^0, PCP and C domain can explain the need to employ such ketosynthases. The experimental results suggest that the length of helix 1 in enacyloxin module 10 ACP and the absence of specific charged residues in its termini may have a negative influence on the propensity to bind the downstream C domain.

Still, many questions remain open and further work will be required to address them.

Chapter 3 addresses the interaction between Bamb_5919 ACP and Bamb_5919 KS^0. To analyse the binding interface of Bamb_5919 ACP, it was important to obtain the backbone assignment, which was then used to analyse ^{15}N relaxation experiments. However, relaxation was only monitored in an unbound protein. It would be also interesting to see how the presence of KS^0 affects the backbone dynamics of ACP. Next, homology model was used as a structure template to solution-state NMR titrations of Bamb_5919 ACP with Bamb_5919 KS^0. Since the NMR data indicate that helix 1 is even longer than predicted from the model, calculating the NMR structure of Bamb_5919 ACP would help to clarify this.

It was emphasized in the text that Bamb_5919 KS^0 is not amenable by solution-state NMR spectroscopy and its crystal structure could not be solved at a good enough resolution. This poses a problem since the interaction interface formed between KS^0 and its carrier protein partners cannot be determined. However, a novel approach called carbene footprinting could overcome this problem[147]. This method uses diazirine-based probe to label the regions of the protein upon irradiation at 349 nm. Areas of the protein complex

that are associated with interactions will remain unlabelled so the interaction interface can be mapped even at the residue level[126].

Finally, mutagenesis of specific residues might also provide the information about the interaction interfaces between enacyloxin chain termination domains.

References

1. Bruegger, J. *et al.* Probing the selectivity and protein×protein interactions of a nonreducing fungal polyketide synthase using mechanism-based crosslinkers. *Chem. Biol.* **20**, 1135–1146 (2013).
2. Anulika, N. P. *et al.* The Chemistry Of Natural Product: Plant Secondary Metabolites. *Int. J. Technol. Enhanc. Emerg. Eng. Res.* **4**, 1 (2016).
3. Dias, D. A. *et al.* Historical overview of natural products in drug discovery. *Metabolites* **2**, 303–336 (2012).
4. Maplestone, R. A. *et al.* The evolutionary role of secondary metabolites - a review. *Gene* **115**, 151–157 (1992).
5. Spellberg, B. *et al.* Trends in Antimicrobial Drug Development: Implications for the Future. *Clin. Infect. Dis.* **38**, 1279–1286 (2004).
6. Lahlou, M. The Success of Natural Products in Drug Discovery. *Pharmacol. & Pharm.* **04**, 17–31 (2013).
7. Mathur, S. & Hoskins, C. Drug development: Lessons from nature. *Biomed. Reports* **6**, 612–614 (2017).
8. Fair, R. J. & Tor, Y. Perspectives in Medicinal Chemistry Antibiotics and Bacterial Resistance in the 21st Century. *Perspect. Medicin. Chem.* 25–64 (2014).
9. Liu, Y. Y. *et al.* Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: A microbiological and molecular biological study. *Lancet Infect. Dis.* **16**, 161–168 (2016).
10. Mahenthiralingam, E. *et al.* The multifarious, multireplicon Burkholderia cepacia complex. *Nat. Rev. Microbiol.* **3**, 144–156 (2005).
11. Mahenthiralingam, E. *et al.* Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic burkholderia ambifaria genomic island. *Chem. Biol.* **18**, 665–677 (2011).
12. Rebic, V. *et al.* The Importance of Acinetobacter Species in the Hospital Environment. *Med. Arch. (Sarajevo, Bosnia Herzegovina)* **72**, 325–329 (2018).
13. Peleg, A. Y. *et al.* Acinetobacter baumannii: Emergence of a successful pathogen.

Clin. Microbiol. Rev. **21**, 538–582 (2008).

14. Berchtold, H. *et al.* Crystal structure of active elongation factor Tu reveals major domain rearrangements. *Nature* **365**, 126–132 (1993).
15. Parmeggiani, A. *et al.* Enacyloxin IIa pinpoints a binding pocket of elongation factor Tu for development of novel antibiotics. *J. Biol. Chem.* **281**, 2893–2900 (2006).
16. Whicher, J.R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. **510**, 560–564 (2014).
17. Khosla, C. Structures and mechanisms of polyketide synthases. *J. Org. Chem.* **74**, 6416–6420 (2009).
18. McAllister, K. A. *et al.* Acyl carrier protein synthases from gram-negative, gram-positive, and atypical bacterial species: Biochemical and structural properties and physiological implications. *J. Bacteriol.* **188**, 4737–4748 (2006).
19. Song, D. *et al.* Alternative method for site-directed mutagenesis of complex polyketide synthase in Streptomyces albus JA3453. *Acta Biochim. Biophys. Sin. (Shanghai).* **40**, 319–326 (2008).
20. Lambalot, R. H. *et al.* A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936 (1996).
21. Elovson J. & Vagelos, P.R. Acyl carrier protein. *J. Biol. Chem.* 3603–3611 (1968).
22. Lambalot, R. H. & Walsh, C. T. Cloning, overproduction, and characterization of the Escherichia coli holo-acyl carrier protein synthase. *J. Biol. Chem.* **270**, 24658–24661 (1995).
23. Copp, J. N. & Neilan, B. A. The phosphopantetheinyl transferase superfamily: Phylogenetic analysis and functional implications in cyanobacteria. *Appl. Environ. Microbiol.* **72**, 2298–2305 (2006).
24. Smith, S. & Tsai, S. C. The type I fatty acid and polyketide synthases: A tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041–1072 (2007).
25. Du, L. & Lou, L. PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **27**, 255–278 (2010).
26. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).
27. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285–295 (2003).

28. Weissman, K. J. Uncovering the structures of modular polyketide synthases. *Nat. Prod. Rep.* **32**, 436–53 (2015).
29. Khosla, C. *et al.* Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annu. Rev. Biochem.* **76**, 195–221 (2007).
30. Caffrey, P. *et al.* Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from Saccharopolyspora erythraea. *FEBS Lett.* **304**, 225–228 (1992).
31. Leadlay, P. F. Enzyme assembly line pictured. *Nature* **510**, 482–483 (2014).
32. Fisch, K. M. *et al.* Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol.* **5**, 494–501 (2009).
33. J. Piel, Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **27**, 996–1047 (2010)
34. Cheng, Y. Q. *et al.* Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3149–3154 (2003).
35. Carreras, C. W. *et al.* Efficient synthesis of aromatic polyketides in vitro by the actinorhodin polyketide synthase. *J. Am. Chem. Soc.* **118**, 5158–5159 (1996).
36. Yu, D. *et al.* Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64**, 285–295 (2012).
37. Abe, I. & Morita, H. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Nat. Prod. Rep.* **27**, 809–838 (2010).
38. Weng, J. K. & Noel, J. P. *Structure-function analyses of plant type III polyketide synthases*. *Methods in Enzymology* **515**, *Elsevier Inc.* (2012)
39. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285–295 (2003).
40. Keatinge-Clay, A. T. The structures of type i polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–1073 (2012).
41. Tang, Y. *et al.* C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124–9 (2006).
42. Pappenberger, G. *et al.* Structure of the Human Fatty Acid Synthase KS-MAT Didomain as a Framework for Inhibitor Design. *J. Mol. Biol.* **397**, 508–519 (2010).
43. Gay, D. C. *et al.* The LINKS motif zippers trans-acyltransferase polyketide synthase

assembly lines into a biosynthetic megacomplex. *J. Struct. Biol.* **193**, 196–205 (2016).

44. Olsen, J. G., *et al.* Structures of β-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. *Structure* **9**, 233–243 (2001).
45. Witkowski, A. *et al.* Conversion of a β-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. *Biochemistry* **38**, 11643–11650 (1999).
46. Wang, J. *et al.* Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature* **441**, 358–361 (2006).
47. Zhang, Y. M. *et al.* Roles of the active site water, histidine 303, and phenylalanine 396 in the catalytic mechanism of the elongation condensing enzyme of streptococcus pneumoniae. *J. Biol. Chem.* **281**, 17390–17399 (2006).
48. Watanabe, K. *et al.* Understanding Substrate Specificity of Polyketide Synthase Modules by Generating Hybrid Multimodular Synthases. *J. Biol. Chem.* **278**, 42020–42026 (2003).
49. Long, P. F. *et al.* Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Mol. Microbiol.* **43**, 1215–1225 (2002).
50. Kosol, S. *et al.* Protein-protein interactions in trans-AT polyketide synthases. *Nat. Prod. Rep.* **35**, 1097–1109 (2018).
51. Huang, Y. *et al.* Characterization of the Ketosynthase and Acyl Carrier Protein Domains at the LnmI Nonribosomal Peptide Synthetase-Polyketide Synthase Interface for Leinamycin Biosynthesis. *Org. Lett.* **18**, 4288–4291 (2016).
52. Parris, K. D. *et al.* Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **8**, 883–895 (2000).
53. Kim, Y. & Prestegard, J. H. Refinement of the NMR structures for acyl carrier protein with scalar coupling data. *Proteins Struct. Funct. Bioinforma.* **8**, 377–385 (1990).
54. Xu, G. Y. *et al.* Solution structure of B. subtilis acyl carrier protein. *Structure* **9**, 277–287 (2001).

55. Holak, T. A. *et al.* Three-Dimensional Structure of Acyl Carrier Protein Determined by NMR Pseudoenergy and Distance Geometry Calculations. *Biochemistry* **27**, 6135–6142 (1988).

56. Smith, S. & Tsai, S. C. The type I fatty acid and polyketide synthases: A tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041–1072 (2007).

57. Chandran, S. S. *et al.* Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains. *Chem. Biol.* **13**, 469–474 (2006).

58. Crosby, J. & Crump, M. P. The structural role of the carrier protein - Active controller or passive carrier. *Nat. Prod. Rep.* **29**, 1111–1137 (2012).

59. Keatinge-Clay, A. T. Stereocontrol within polyketide assembly lines. *Nat. Prod. Rep.* **33**, 141–149 (2016).

60. Piasecki, S. K. *et al.* Employing modular polyketide synthase ketoreductases as biocatalysts in the preparative chemoenzymatic syntheses of diketide chiral building blocks. *Chem. Biol.* **18**, 1331–1340 (2011).

61. Noey, E. L. *et al.* Origins of stereoselectivity in evolved ketoreductases. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E7065–E7072 (2015).

62. Tang, Y. *et al.* The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124–11129 (2006).

63. Fiers, W. D. *et al.* Vinylogous Dehydration by a Polyketide Dehydratase Domain in Curacin Biosynthesis. *J. Am. Chem. Soc.* **138**, 16024–16036 (2016).

64. Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **384**, 941–953 (2008).

65. Broadhurst, R.W. *et al.* The Structure of Docking Domains in Modular Polyketide Synthases. *J. Chem. Biol.*, **10**, 723-731 (2003)

66. Tsai, S. C. *et al.* Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: Versatility from a unique substrate channel. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14808–14813 (2001).

67. Buchholz, T. J. *et al.* Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase docking Domains. *ACS Chem Biol.* **4**, 41- 52 (2009)

68. Zheng, J. *et al.* The missing linker: a dimerization motif located within polyketide

synthase modules, *ACS Chem Biol.* **8**, 1263-1270 (2013)

69. Brink, J. *et al.* Quaternary structure of human fatty acid synthase by electron cryomicroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 138–143 (2002).
70. Maier, T. *et al.* N. Architecture of mammalian fatty acid synthase at 4.5Å resolution. *Science* **311**, 1258–1262 (2006).
71. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–7 (2014).
72. Challis, G. L. & Naismith, J. H. Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol.* **14**, 748–756 (2004).
73. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angew. Chemie - Int. Ed.* **56**, 3770–3821 (2017).
74. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).
75. Wang, L., Yuan, M. & Zheng, J. Crystal structure of the condensation domain from lovastatin polyketide synthase. *Synth. Syst. Biotechnol.* **4**, 10–15 (2019).
76. Samel, S. A. *et al.* Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure* **15**, 781–792 (2007).
77. Zhang, J. *et al.* Structural basis of nonribosomal peptide macrocyclization in fungi. *Nat. Chem. Biol.* **12**, 1001–1003 (2016).
78. Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238 (2016).
79. Bloudoff, K. *et al.* Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chem. Biol.* **23**, 331–339 (2016).
80. Keating, T. A. *et al.* The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–526 (2002).
81. Goodrich, A. C. *et al.* Solution Structure of a Nonribosomal Peptide Synthetase Carrier Protein Loaded with Its Substrate Reveals Transient, Well-Defined Contacts. *J. Am. Chem. Soc.* **137**, 12100–12109 (2015).
82. Jaremko, M. J. *et al.* Structure and Substrate Sequestration in the Pyoluteorin Type

II Peptidyl Carrier Protein PltL. *J. Am. Chem. Soc.* **137**, 11546–11549 (2015).

83. Beld, J. *et al.* Visualizing the chain-flipping mechanism in fatty-acid biosynthesis. *Angew. Chemie - Int. Ed.* **53**, 14456–14461 (2014).
84. Kittilä, T. *et al.* New Structural Data Reveal the Motion of Carrier Proteins in Nonribosomal Peptide Synthesis. *Angew. Chemie - Int. Ed.* **55**, 9834–9840 (2016).
85. Barajas, J. F. *et al.* Engineered polyketides: Synergy between protein and host level engineering. *Synth. Syst. Biotechnol.* **2**, 147–166 (2017).
86. Oliynyk, M. *et al.* A hybrid modular polyketide synthase obtained by domain swapping. *Chem. Biol.* **3**, 833–839 (1996).
87. Masschelein, J. *et al.* A dual transacylation mechanism for polyketide synthase chain release in enacyloxin antibiotic biosynthesis, *Nat. Chem.* **11**, 906-912 (2019)
88. Raran-Kurussi, S. *et al.* Differential temperature dependence of tobacco etch virus and rhinovirus 3C proteases. *Anal. Biochem.* **436**, 142–144 (2013).
89. Keller R, Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment, PhD thesis, ETH (2004)
90. Kelley, L. A. *et al.* The Phyre2 web portal for protein modeling , prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
91. Alekseyev, V. Y. *et al.* Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci* **16**, 2093–2107 (2007).
92. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
93. Sreerama, N. *et al.* Estimation of the number of α-helical and β-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci.* **8**, 370–380 (2008).
94. Keller, S. *et al.* High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.* **84**, 5066–5073 (2012).
95. Keller, R. Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment. ETH Zürich (2004)
96. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127**, 8014–8015 (2005).
97. Lescop, E. *et al.* A set of BEST triple-resonance experiments for time-optimized

protein resonance assignment. *J. Magn. Reson.* **187**, 163–169 (2007).

98. Grzesiek, S. & Bax, A. Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. *J. Biomol. NMR* **3**, 185–204 (1993).

99. Schanda, P. *et al.* Speeding up three-dimensional protein NMR experiments to a few minutes. *J. Am. Chem. Soc.* **128**, 9042–9043 (2006).

100. Grzesiek, S. & Bax, A. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Magn. Reson.* **96**, 432–440 (1992).

101. Schleucher, J. *et al.* A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR* **4**, 301–306 (1994).

102. Kay, L. E. *et al.* Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation, *J. Magn. Reson.* **109**, 129-133 (1994)

103. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99**, 201–207 (1992).

104. Kay, L. E. *et al.* A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlations in H_2O samples in proteins, *J. Magn. Reson.*, **101**, 333-337 (1993)

105. Palmer, A. G. *et al.* Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson.* **93**, 151–170 (1991).

106. Kay, L. E., Keifer, P. & Saarinen, T. Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665 (1992).

107. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).

108. Zhou, D. H. & Rienstra, C. M. High-performance solvent suppression for proton detected solid-state NMR. *J. Magn. Reson.* **192**, 167–172 (2008).

109. Kelley, L.A. *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845-858 (2015)

110. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).

111. Dougherty, W. G. *et al.* Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO J.* **7**, 1281–1287 (1988).

112. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890 (2007).

113. Banerjee, B. *et al.* Chapter 2. Circular dichroism. Data Processing Handbook for Complex Biological Data Sources. Elsevier Inc. (2019)

114. Zornetzer, G. A. *et al.* The length of the bound fatty acid influences the dynamics of the acyl carrier protein and the stability of the thioester bond. *Biochemistry* **49**, 470–477 (2010).

115. Shah, N. B. & Duncan, T. M. Bio-layer interferometry for measuring kinetics of protein-protein interactions and allosteric ligand effects. *J. Vis. Exp.* 1–7 (2014). doi:10.3791/51383

116. Nirschl, M. *et al.* Review of transducer principles for label-free biomolecular interaction analysis. *Biosensors* **1**, 70–92 (2011).

117. Meyerkord, C. L. & Fu, H. Protein-protein interactions: Methods and applications: Second edition. *Protein-Protein Interact. Methods Appl. Second Ed.* **1278**, 1–613 (2015).

118. Moretto, L. *et al.* Dissecting how modular polyketide synthase ketoreductases interact with acyl carrier protein-attached substrates. *Chem. Commun.* **53**, 11457–11460 (2017).

119. Pervushin, K. *et al.* Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12366–71 (1997).

120. Zhang, H. & van Ingen, H. Isotope-labeling strategies for solution NMR studies of macromolecular assemblies. *Curr. Opin. Struct. Biol.* **38**, 75–82 (2016).

121. Clore, G. M. *et al.* Analysis of the Backbone Dynamics of Interleukin-1-β Using 2-Dimensional Inverse Detected Heteronuclear ^{15}N-^{1}H NMR-Spectroscopy. *Biochemistry* **29**, 7387-7401 (1990)

122. Copie, V. *et al.* Secondary structure of β-hydroxydecanoyl thiol ester dehydrase, a 39-kDa protein, derived from Hα, Cα, Cβ and CO signal assignments and the Chemical Shift Index: comparison with the crystal structure. *J Biomol NMR*, 7, 335-

340 (1996)

123. Kay, L.E. *et al.* Backbone Dynamics of Proteins as Studied by ^{15}N Inverse Detected Heteronuclear NMR-Spectroscopy - Application to Staphylococcal Nuclease. *Biochemistry*, **28**, 8972-8979 (1989)

124. Wishart, D.S. et al. The Chemical Shift Index: A Fast and Simple Method for the Assignment of Protein Secondary Structure through NMR Spectroscopy. *Biochemistry*, **31**, 1647-1651 (1992)

125. Kovermann, M. *et al.* Protein dynamics and function from solution state NMR spectroscopy. *Q. Rev. Biophys.* **49**, e6 (2016).

126. Keeler, J. Understanding NMR Spectroscopy. *John Wiley Sons* 409–416 (2010).

127. Clore, G. M. *et al.* P. T. Analysis of the Backbone Dynamics of Interleukin-1β Using Two-Dimensional Inverse Detected Heteronuclear 15N-1H NMR Spectroscopy. *Biochemistry* **29**, 7387–7401 (1990).

128. Kay, L. E. *et al.* Backbone Dynamics of Proteins As Studied by 15N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* **28**, 8972–8979 (1989).

129. Copié, V. *et al.* Secondary structure of β-hydroxydecanoyl thiol ester dehydrase, a 39-kDa protein, derived from Hα, Cα, Cβ and CO signal assignments and the Chemical Shift Index: Comparison with the crystal structure. *J. Biomol. NMR* **7**, 335–340 (1996).

130. Lim, J. *et al.* Solution structures of the Acyl carrier protein domain from the highly reducing type I iterative polyketide synthase CalE8. *PLoS One* **6**, 1–12 (2011).

131. Flaman, A. S. *et al.* Site-directed Mutagenesis of Acyl Carrier Protein (ACP) Reveals Amino Acid Residues Involved in ACP Structure and Acyl-ACP Synthetase Activity. *J. Biol. Chem.* **276**, 35934–35939 (2001).

132. Worsham, L. M. S. *et al.* Amino acid residues of Escherichia coli acyl carrier protein involved in heterologous protein interactions. *Biochemistry* **42**, 167–176 (2003).

133. Zhang, Y. M. *et al.* Identification and Analysis of the Acyl Carrier Protein (ACP) Docking Site on β-Ketoacyl-ACP Synthase III. *J. Biol. Chem.* **276**, 8231–8238 (2001).

134. Zhang, Y. M. *et al.* The application of computational methods to explore the diversity and structure of bacterial fatty acid synthase. *J. Lipid Res.* **44**, 1–10 (2003).

135. Milligan, J. C. *et al.* Molecular basis for interactions between an acyl carrier protein

and a ketosynthase. *Nat. Chem. Biol.* **15**, 669–671 (2019).

136. Beltran-Alvarez, P. *et al.* Preliminary kinetic analysis of acyl carrier protein-ketoacylsynthase interactions in the actinorhodin minimal polyketide synthase. *Mol. Biosyst.* **5**, 511–518 (2009).
137. Ye, Z. & Williams, G. J. Mapping a ketosynthase: Acyl carrier protein binding interface via unnatural amino acid-mediated photo-cross-linking. *Biochemistry* **53**, 7494–7502 (2014).
138. Weissman, K. J. *et al.* Evidence for a Protein-Protein Interaction Motif on an Acyl Carrier Protein Domain from a Modular Polyketide Synthase. *Chem. Biol.* **13**, 625–636 (2006).
139. Kapur, S. *et al.* Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22066–22071 (2010).
140. Kapur, S. *et al.* Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4110–4115 (2012).
141. Charkoudian, L. K. *et al.* Probing the interactions of an acyl carrier protein domain from the 6-deoxyerythronolide B synthase. **20**, 1244–1255 (2011).
142. Milligan, J.C. Synthetic Substrate Analogues as Tools for Structural Studies of Protein-Protein and Protein- Substrate Interactions in Microbial Biosynthesis, UC Irvine (2017)
143. Tycko, R. Solid-State NMR Studies of Amyloid Fibril Structure. *Annu. Rev. Phys. Chem.* **62**, 279–299 (2011).
144. Ladizhansky, V. Applications of solid-state NMR to membrane proteins. *Biochim. Biophys. Acta - Proteins Proteomics* **1865**, 1577–1586 (2017).
145. Martin, R. W. *et al.* Advances in instrumentation and methodology for solid-state NMR of biological assemblies. *J. Struct. Biol.* **206**, 73–89 (2019).
146. Mainz, A. *et al.* Large protein complexes with extreme rotational correlation times investigated in solution by magic-angle-spinning NMR spectroscopy. *J. Am. Chem. Soc.* **131**, 15968–15969 (2009).
147. Kosol, S. *et al.* Molecular basis for condensation domain-mediated chain release from the enacyloxin polyketide synthase. *Nat. Chem., resubmitted after Revis.* (2018).

148. Chevelkov, V. *et al.* Differential line broadening in MAS solid-state NMR due to dynamic interference. *J. Am. Chem. Soc.* **129**, 10195–10200 (2007).

149. Leo, G. C. *et al.* Dynamics of fd Coat Protein in Lipid Bilayers. *Biochemistry* **26**, 854–862 (1987).

150. Chevelkov, V. *et al.* Comparison of solid-state dipolar couplings and solution relaxation data provides insight into protein backbone dynamics. *J. Am. Chem. Soc.* **132**, 5015–5017 (2010).

151. Franks, W. T. *et al.* Magic-angle spinning solid-state NMR spectroscopy of the β1 immunoglobulin binding domain of protein G (GB1): ^{15}N and ^{13}C chemical shift assignments and conformational analysis. *J. Am. Chem. Soc.* **127**, 12291–12305 (2005).

152. Jardetzky, O. & Wade-Jardetzky, N. G. Comparison of protein structures by high resolution solid state and solution NMR. *FEBS Lett.* **110**, 133–135 (1980).

153. Agarwal, V. *et al.* Protein side-chain dynamics as observed by solution- and solid-state NMR spectroscopy: A similarity revealed. *J. Am. Chem. Soc.* **130**, 16611–16621 (2008).

154. Drake E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238 (2016).

155. Koglin, A. *et al.* Conformational switches modulate protein interactions in peptide antibiotic synthetases, *Sci. Rep.* **312**, 273-276 (2006)

156. Goodrich, A. C. & Frueh, D. P. A nuclear magnetic resonance method for probing molecular influences of substrate loading in nonribosomal peptide synthetase carrier proteins. *Biochemistry* **54**, 1154–1156 (2015).

157. Weber, T. *et al.* Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* **8**, 407–418 (2000).

158. Gulick, A. M. Structural insight into the necessary conformational changes of modular nonribosomal peptide synthetases. *Curr. Opin. Chem. Biol.* **35**, 89–96 (2016).

159. Miller, B. R. *et al.* Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. *J. Biol. Chem.* **291**, 22559–22571 (2016).

160. Frueh, D. P. *et al.* Dynamic thiolation-thioesterase structure of a non-ribosomal

peptide synthetase. *Nature* **454**, 903–906 (2008).

161. Koglin, A. *et al.* Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* **454**, 907–911 (2008).
162. Dinkel, H. *et al.* The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* **42**, (2014).
163. Fuxreiter, M. *et al.* Local structural disorder imparts plasticity on linear motifs. *Bioinformatics* **23**, 950–956 (2007).
164. Davey, N. E. *et al.* Attributes of short linear motifs. *Mol. Biosyst.* **8**, 268–281 (2012).
165. Diella, F. *et al.* Understanding eukaryotic linear motifs and their role in cell signalling and regulation, Front. Biosci **13**, 6580-6603 (2008).
166. Broadhurst, R.W. *et al.* The structure of docking domains in modular polyketide synthases, *Chem. Biol.* **10**, 723-731 (2003)
167. Weissman, K. J. The structural basis for docking in modular polyketide biosynthesis. *ChemBioChem* **7**, 485–494 (2006).
168. Tsuji, S. Y. *et al.* Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry* **40**, 2326–2331 (2001).
169. Wu, N., Cane *et al.* Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry* **41**, 5056–5066 (2002).

Supplementary

Constructs

His-tag

MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT

Bam_5919 ACP (shorter construct) (from G1626 to P1669)

[MW (+His-tag): 17994.37 Da, MW(+GIDPFT): 14841.83 Da]

GAPRVERRAGGTAGAALLAGLASLDAAARAARLKRHLEAAIRKLLNRADTLDDRASMFDLGLDSLLS
IDLMQLEKDLACSLSTTVLHDHPTIEALAGFLAERVGAPPAGTVRAGAAGGAGAGTGAPAGATGA
AAAHAVSSASP

Bamb_5919 ACP(longer construct) (from L1614 to A1800)

[MW (+His-tag): 21662.36 Da, MW(+GIDPFT): 18509 Da]

LAELVEAEAGATGASGAPRVERRAGGTAGAALLAGLASLDAAARAARLKRHLEAAIRKLLNRADTLD
DRASMFDLGLDSLLSIDLRMQLEKDLACSLSTTVLHDHPTIEALAGFLAERVGAPPAGTVRAGAAGG
AGAGTGAPAGATGAAAAHAVSSASPVPAGAASAAASAASAAAAAGA
PSRATFA

Bamb_5919 KS^0 (from P1794 to A2463)

[MW (+His-tag): 74956.31 Da]

PSRATFAAEPRRAGGAALPPGAGPDDIAIIGVSGRYPGAADLGAFWDNLRDGHDAITPIPPERWNH
DAYFDRQRNVPGKSYSAWGGFIEDVDAFDPAFFSISPRMSAYLDPKERLFLETVWNLLEEAGETRER
MQQAYGAQVGVFVGAMYQLYGACAADEGERVATALSSYNAIAHRTSYFFNLRGPSIALDTMCSSS
LTAVHYACRSLLDGDCALAIAGGVNLSLHPRKYVGLSQAQIVGSHADSRSFSDGDGYLPAEGVGAVL
LKPLARALADDDRILAVIKASSVNHGGRATGYYAPNANAQVDLMEASFRKAGVSPESIDYIEAAANG
TSLGDAVELRALARVFDGTARDGRRVPIGTVKSNIGHPEAASGIAQLTKVILQMQHETLVPSIKTEPV
NPNLDLAHTPFRLLSRQAAWPSDPARPRRATVSSFGASGANAHLIVEAFETVEAEPAPAVAQAAPP
AEIVVLSARTPAQLREVARRLLAWLATRQAAGRAESAVPLAERGRACSLANLAHTLQIGREAMDCR
LALLADSLDTLGDGLRRFLGESAGAAEPAIYHGNVQDQLEMRNLLAGAAGDAMAQTLVAERNLEG
LMLHWVQGGNVPWAALREGRPARRLVLPTYPFERERYWLSGASDAAGRGAGEPQVPAEPA
EAASEPSVVDGRA

Bamb_5917 PCP (from G215 to the end – R315)

[MW (+His-tag): 14466.32 Da, MW(+GIDPFT): 18509.82 Da]

GAAAGVSAAGIEPDLTAIWQALFALPAVGRHQDFFALGGDSQLGLRMLAQLRERHGVDLPLRCLY
EAPTVARLAETIVRLAAPAPSGDQDDASEYEEGVIR

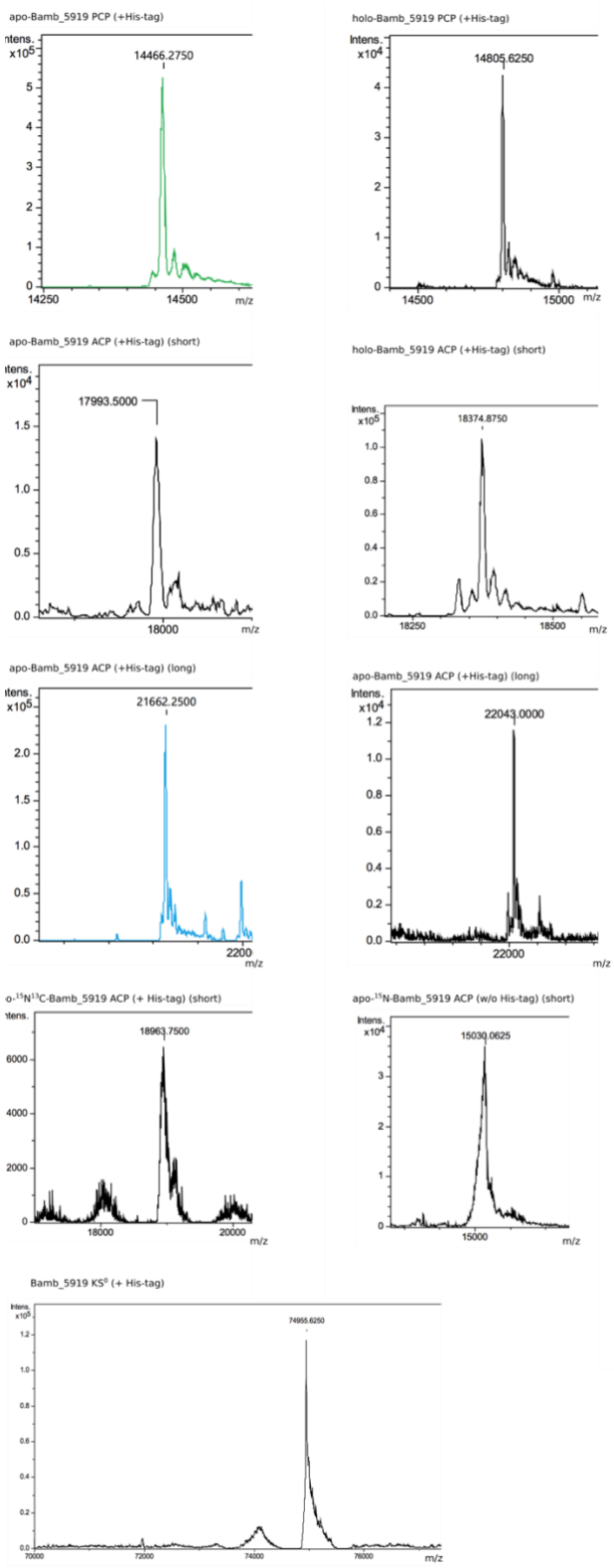


Figure S.2 LC-MS analysis of used protein constructs. The conversion of apo to holo protein adds 358 Da (Ppant arm) to the mass of apo protein.

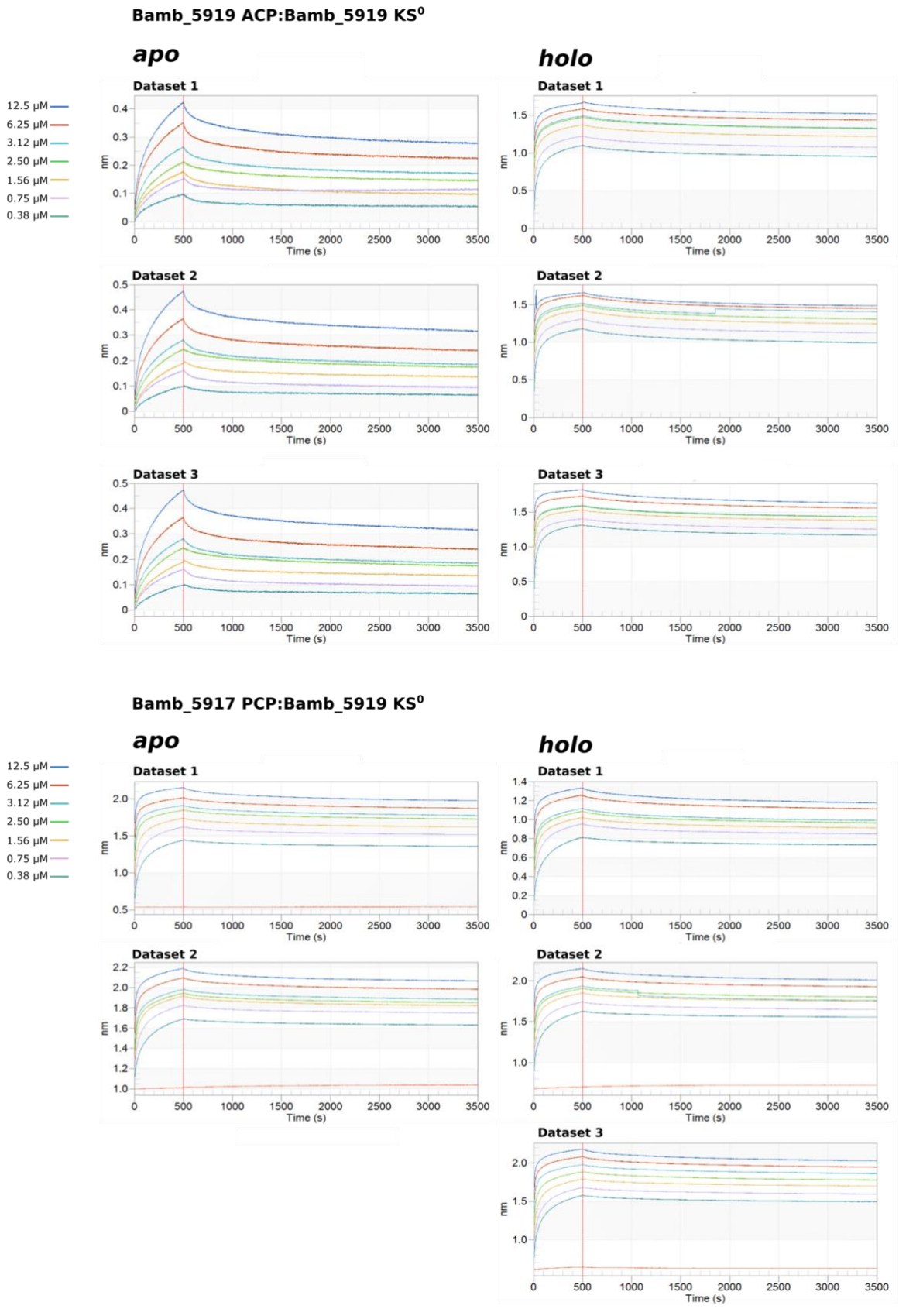


Figure S.3 BLI – binding curves

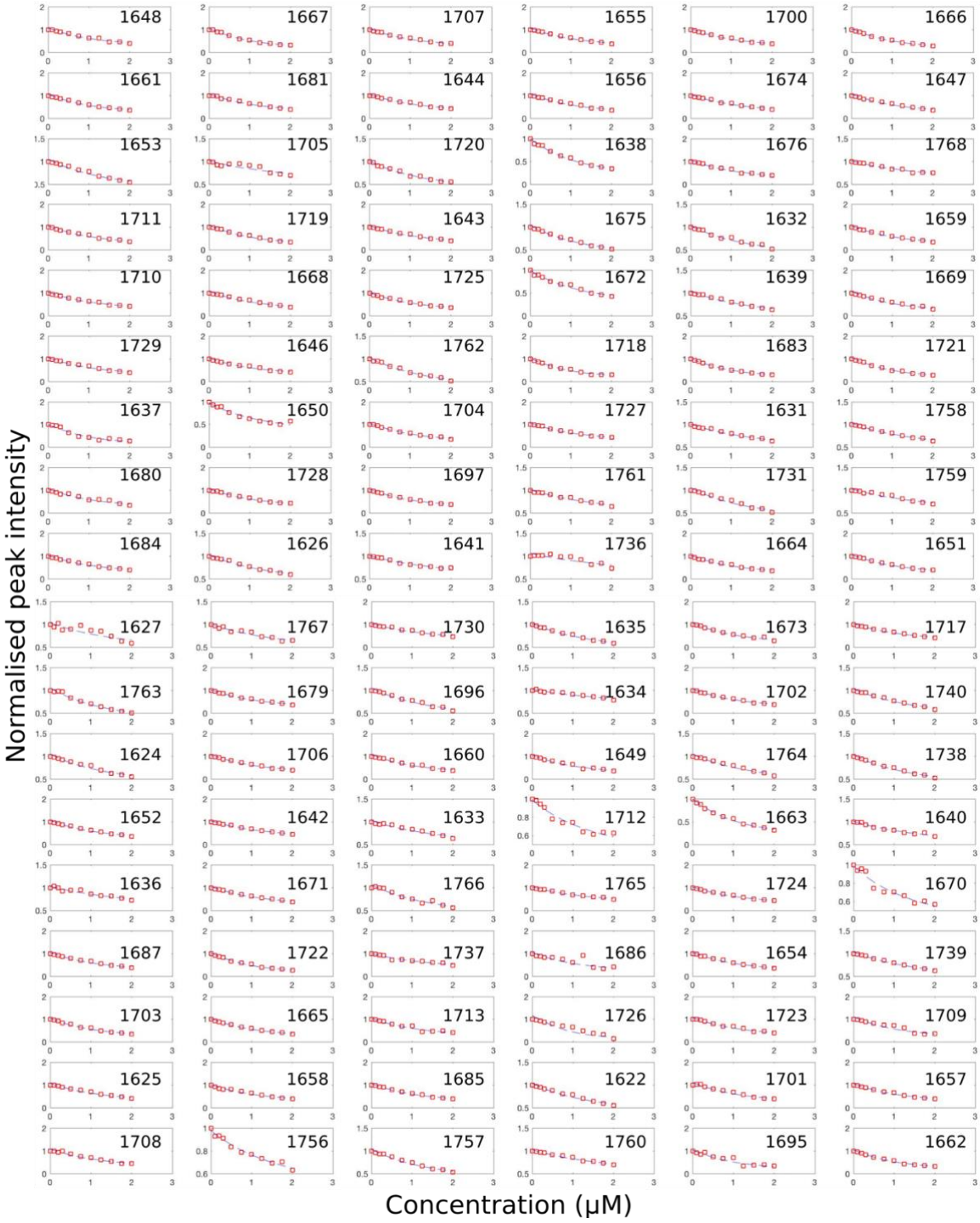


Figure S.4-1 Dataset 1 - titrations profiles for the peaks disappearing in the NMR titration of the ^{15}N-apo-Bamb_5917 ACP with Bamb_5919 KS^0. The squares represent peak volumes normalised to the peak volume of the apo-Bamb-5919 ACP before addition of Bamb_5919 KS^0. The lines represent the best fit curves to equation $[P] = \frac{A\left(-K_d-[L]_t+[P]_t+\sqrt{4K_d[P]_t+(K_d+[L]_t-[P]_t)^2}\right)}{2[P]_t}$.

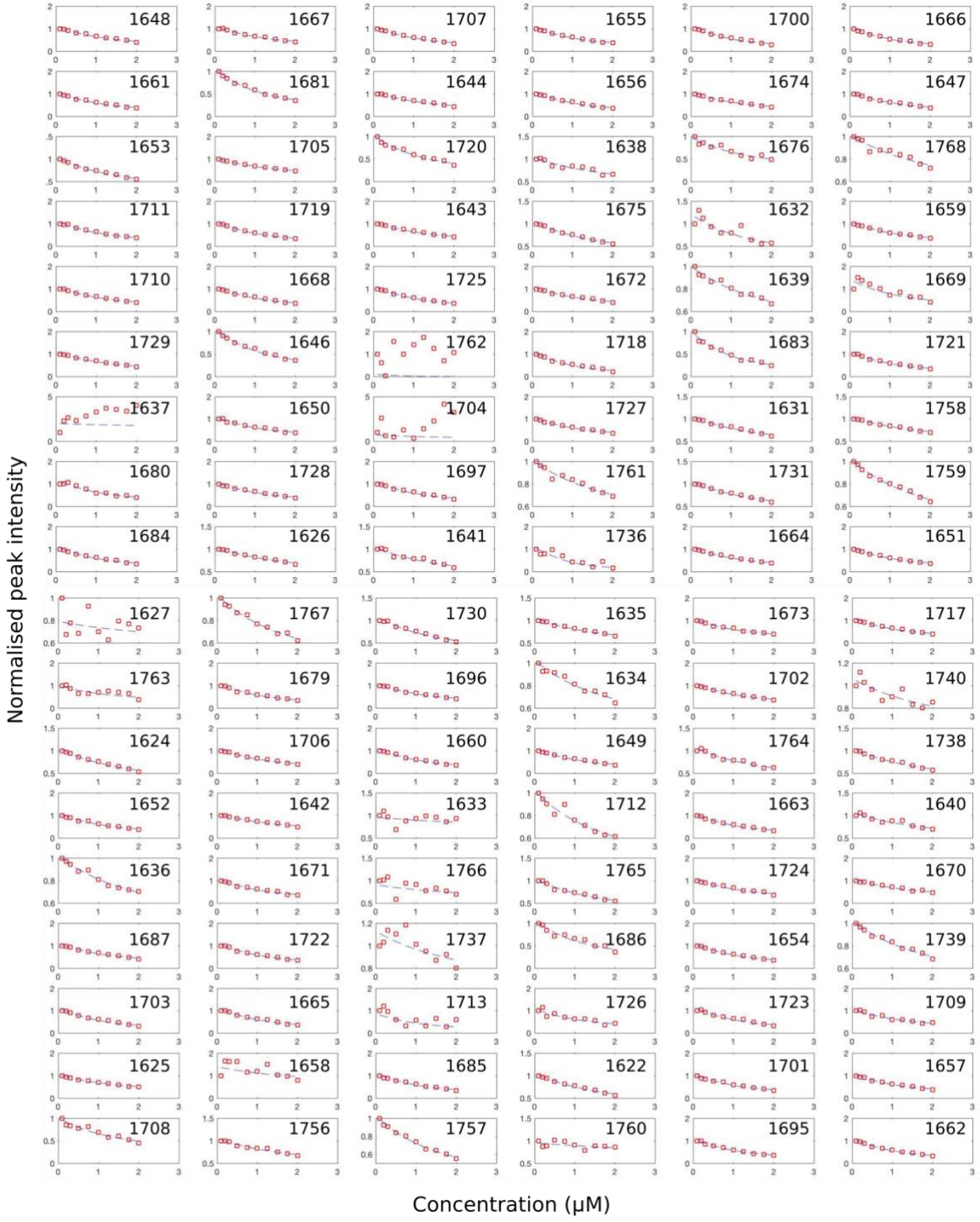


Figure S.4-2 Dataset 2 - titrations profiles for the peaks disappearing in the NMR titration of the ^{15}N-apo-Bamb_5917 ACP with Bamb_5919 KS^{0}

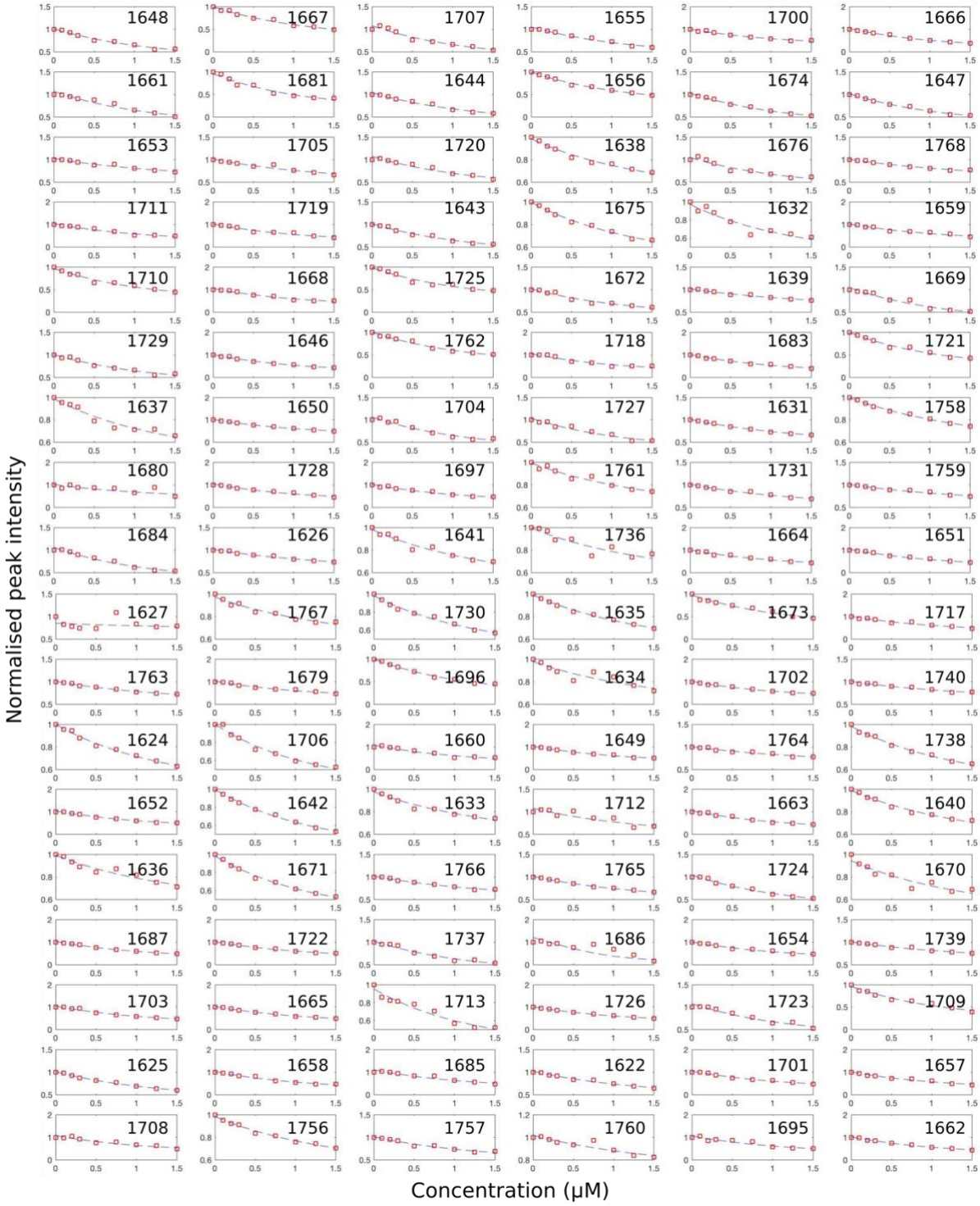


Figure S.4-3 Dataset 3 - titrations profiles for the peaks disappearing in the NMR titration of the ^{15}N-apo-Bamb_5917 ACP with Bamb_5919 KS^0

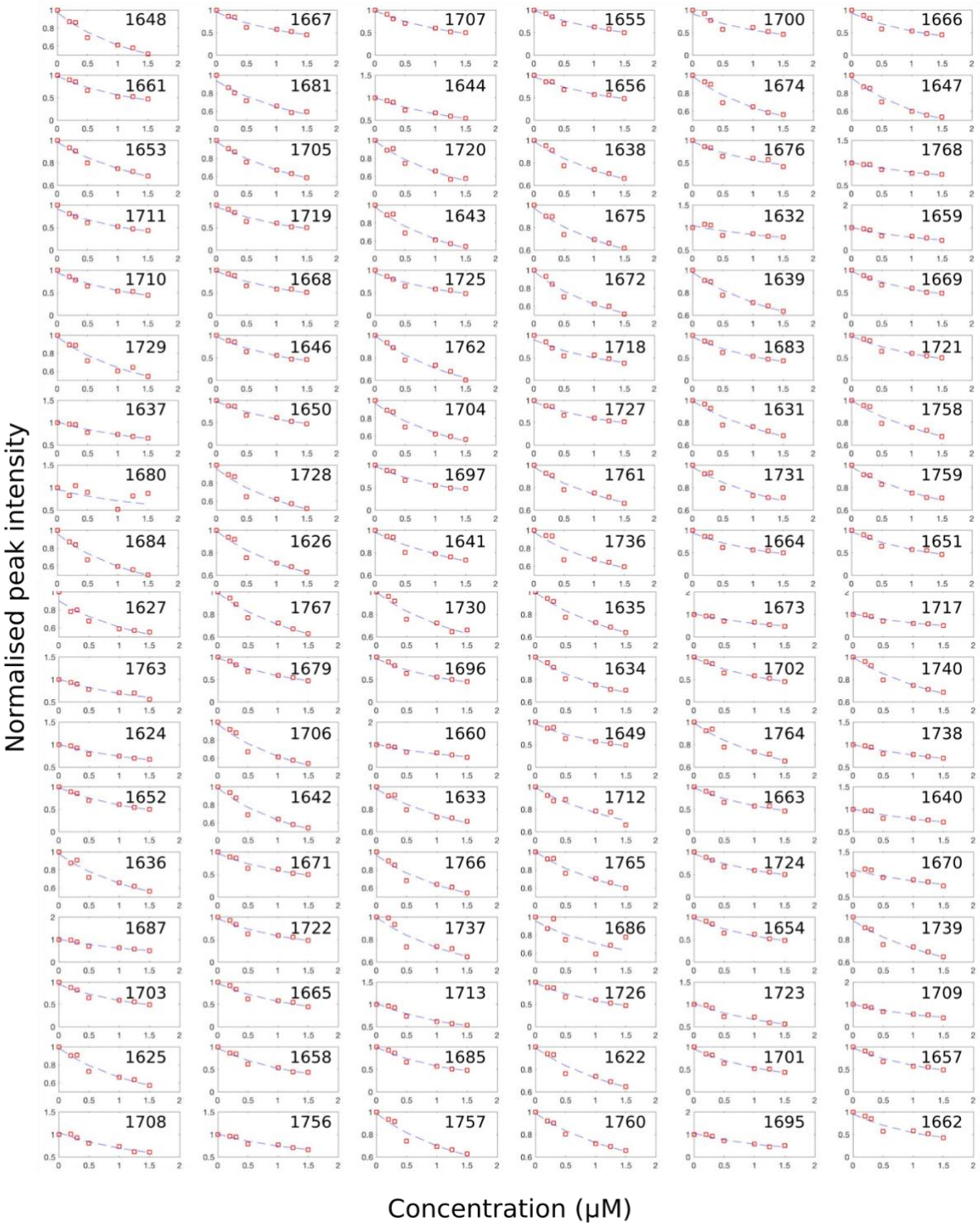


Figure S.4-4 Dataset 4 - titrations profiles for the peaks disappearing in the NMR titration of the ^{15}N-apo-Bamb_5917 ACP with Bamb_5919 KS^0

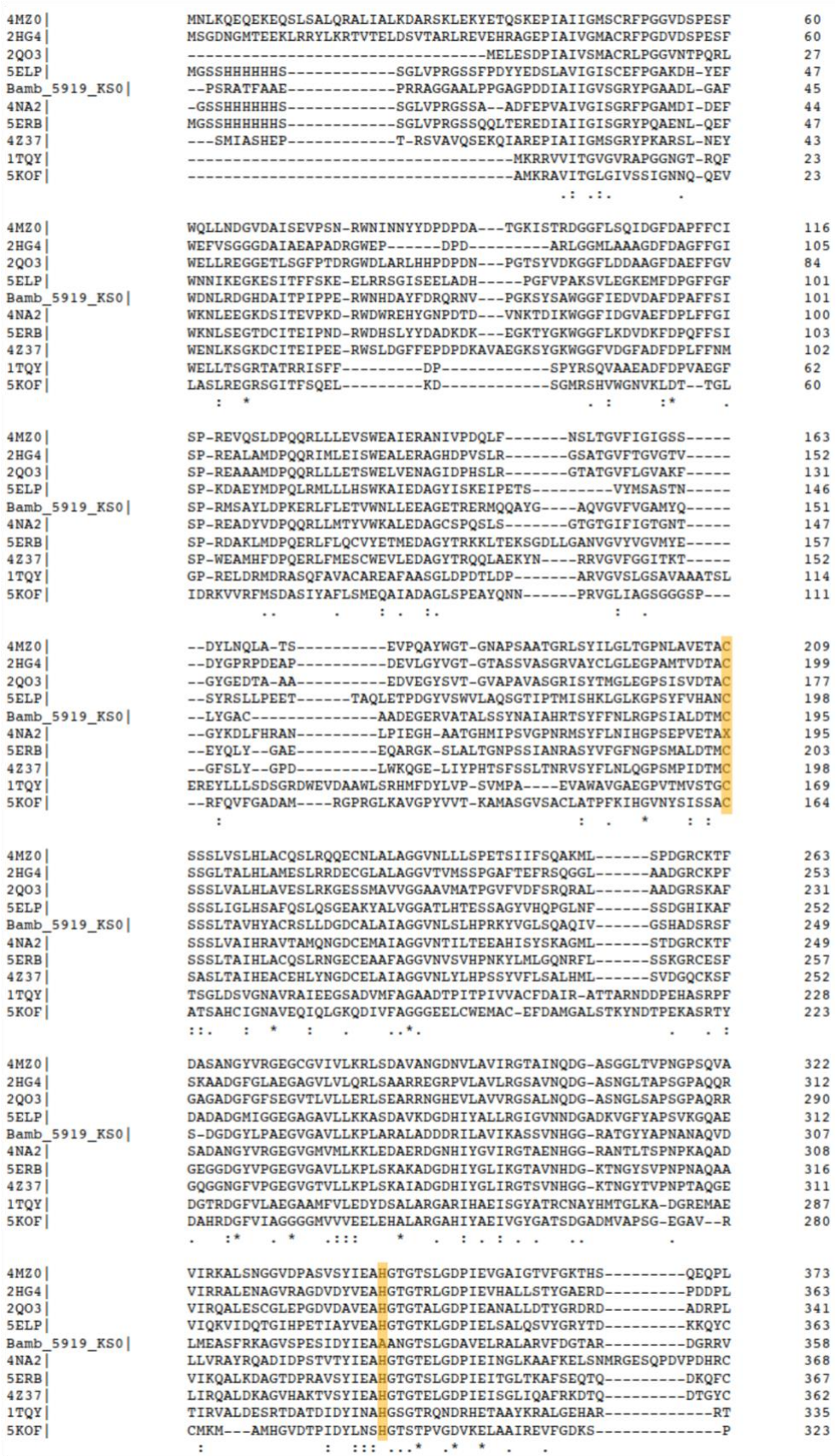
Table S.1 Chemical shift assignments of apo-Bamb_5919 ACP

Amino acid	#	H (ppm)	N (ppm)	Cα (ppm)	Cβ (ppm)
D	tag	8.581	127.061	51.961	41.759
P	tag	-	-	-	-
F	tag	-	-	-	-
T	tag	7.971	112.861	62.032	69.754
G	1626	8.484	111.164	45.024	-
A	1627	8.0	123.532	50.301	-
P	1628	-	-	-	-
R	1629	-	-	-	-
V	1630	-	-	-	-
E	1631	8.569	125.713	56.203	32.716
R	1632	8.485	122.746	62.101	32.716
R	1633	8.485	125.746	55.799	43.174
A	1634	8.484	126.042	52.500	19.281
G	1635	8.466	108.220	45.269	-
G	1636	8.358	108.525	45.304	-
T	1637	8.125	112.679	61.848	69.700
A	1638	8.516	126.298	53.095	19.094
G	1639	8.384	108.168	45.164	-
A	1640	8.323	123.749	52.679	19.067
A	1641	8.133	122.415	54.095	19.052
L	1642	8.155	120.509	57.485	-
L	1643	7.917	116.413	56.309	42.208
A	1644	7.704	121.969	52.275	18.125
G	1645	-	-	-	-
L	1646	-	-	-	-
A	1647	-	-	-	-
S	1648	7.655	110.124	58.786	-
L	1649	7.324	121.857	53.799	37.269
D	1650	8.273	119.649	53.099	-
A	1651	8.635	121.641	56.718	18.240
A	1652	8.379	120.783	55.081	17.678
A	1653	8.252	123.531	54.612	-
R	1654	9.043	119.956	60.106	35.959
A	1655	8.141	120.765	55.671	17.474
A	1656	7.872	119.699	55.557	35.448
R	1657	7.925	118.304	58.356	24.732
L	1658	8.5	120.427	57.97	-
K	1659	8.814	121.318	61.385	21.259
R	1660	7.262	116.907	58.981	-
H	1661	8.035	120.508	60.435	17.242
L	1662	8.909	118.499	57.648	41.520

E	1663	8.535	118.853	60.857	43.334
A	1664	7.484	119.130	55.094	-
A	1665	-	-	-	-
I	1666	8.577	120.840	61.398	-
R	1667	7.513	116.6	59.058	43.195
K	1668	7.778	118.290	59.944	37.813
L	1669	7.622	122.407	56.726	-
L	1670	7.741	117.507	55.932	42.349
N	1671	-	-	-	-
R	1672	7.94	116.102	54.710	17.781
A	1673	8.756	124.104	52.163	-
D	1674	8.514	118.004	52.862	41.112
T	1675	7.988	114.130	62.810	50.396
L	1676	8.919	126.635	53.685	43.398
D	1677	7.974	117.770	54.144	42.193
D	1678	8.598	125.966	55.809	-
R	1679	8.487	115.924	55.427	-
A	1680	8.943	116.092	57.298	63.543
S	1681	-	-	-	-
M	1682	9.823	126.621	52.862	40.961
F	1683	9.020	121.037	57.512	35.969
D	1684	-	-	-	-
L	1685	-	-	-	-
G	1686	-	-	-	-
L	1687	-	-	-	-
D	1688	-	-	-	-
S	1689	-	-	-	-
L	1690	-	-	-	-
L	1691	-	-	-	-
S	1692	-	-	-	-
I	1693	-	-	-	-
D	1694	-	-	-	-
L	1695	-	-	-	-
R	1696	8.422	119.637	58.007	-
M	1697	8.449	115.423	58.695	33.236
Q	1698	8.185	120.218	58.790	-
L	1699	8.401	118.601	57.88	-
E	1700	8.248	117.790	59.987	35.948
K	1701	7.435	118.16	58.888	43.465
D	1702	8.962	120.278	57.419	36.003
L	1703	8.254	114.417	53.908	-
A	1704	7.618	122.332	52.396	19.283
C	1705	8.096	112.508	55.666	43.032
S	1706	8.171	114.311	57.298	63.444

L	1707	-	-	-	-
S	1708	-	-	-	-
T	1709	7.332	120.137	57.770	41.490
T	1710	8.393	113.543	62.544	20.875
V	1711	7.407	122.527	65.931	31.004
L	1712	8.060	118.279	56.720	-
H	1713	7.335	115.806	58.881	43.452
D	1714	8.134	117.721	55.679	30.563
H	1715	7.102	116.630	52.633	40.947
P	1716	-	-	-	-
T	1717	-	-	-	-
I	1718	-	-	-	-
E	1719	-	-	-	-
A	1720	-	-	-	-
L	1721	-	-	-	-
A	1722	-	-	-	-
G	1723	-	-	-	-
F	1724	7.818	122.403	60.572	38.883
L	1725	8.283	118.282	57.085	42.166
A	1726	8.283	119.645	55.199	18.533
E	1727	7.164	115.665	57.532	-
R	1728	7.572	118.131	56.486	31.278
V	1729	7.601	115.242	62.611	-
G	1730	7.935	109.186	44.891	-
A	1731	7.842	123.952	50.343	18.026
P	1732	-	-	-	-
P	1733	-	-	-	-
A	1734	-	-	-	-
G	1735	-	-	-	-
T	1736	-	-	-	-
V	1737	8.248	123.236	62.321	32.706
R	1738	7.690	119.921	57.527	40.437
A	1739	8.032	114.017	54.330	18.739
G	1740	8.455	108.167	45.558	-
A	1741	8.155	123.75	53.106	19.385
A	1742	8.268	123.094	53.866	37.184
G	1743	8.245	108.483	45.153	-
G	1744				
A	1745	8.207	123.785	52.630	18.2
G	1746	8.179	113.194	46.2	-
A	1747				
G	1748				
T	1749	8.025	114.018	62.144	32.637
G	1750	8.411	110.943	45.272	-
A	1751	8.095	124.692	50.287	17.990

P	1752	-	-	-	-
A	1753	-	-	-	-
G	1754	-	-	-	-
A	1755	-	-	-	-
T	1756	-	-	-	-
G	1757	-	-	-	-
A	1758	-	-	-	-
A	1759	8.419	125.999	52.666	19.236
A	1760	8.264	122.818	52.498	19.064
A	1761	8.134	122.884	52.424	18.895
H	1762	8.130	118.220	55.786	30.642
A	1763	-	-	-	-
V	1764	8.657	119.276	55.786	30.642
S	1765	7.133	109.415	57.994	63.647
S	1766	-	-	-	-
A	1767	8.292	125.806	52.397	19.694
S	1768	8.257	117.006	56.313	63.553
P	1769	-	-	-	-


4MZ0|
2HG4|
2QO3|
5ELP|
Bamb_5919_KS0|
4NA2|
5ERB|
4Z37|
1TQY|
5KOF|

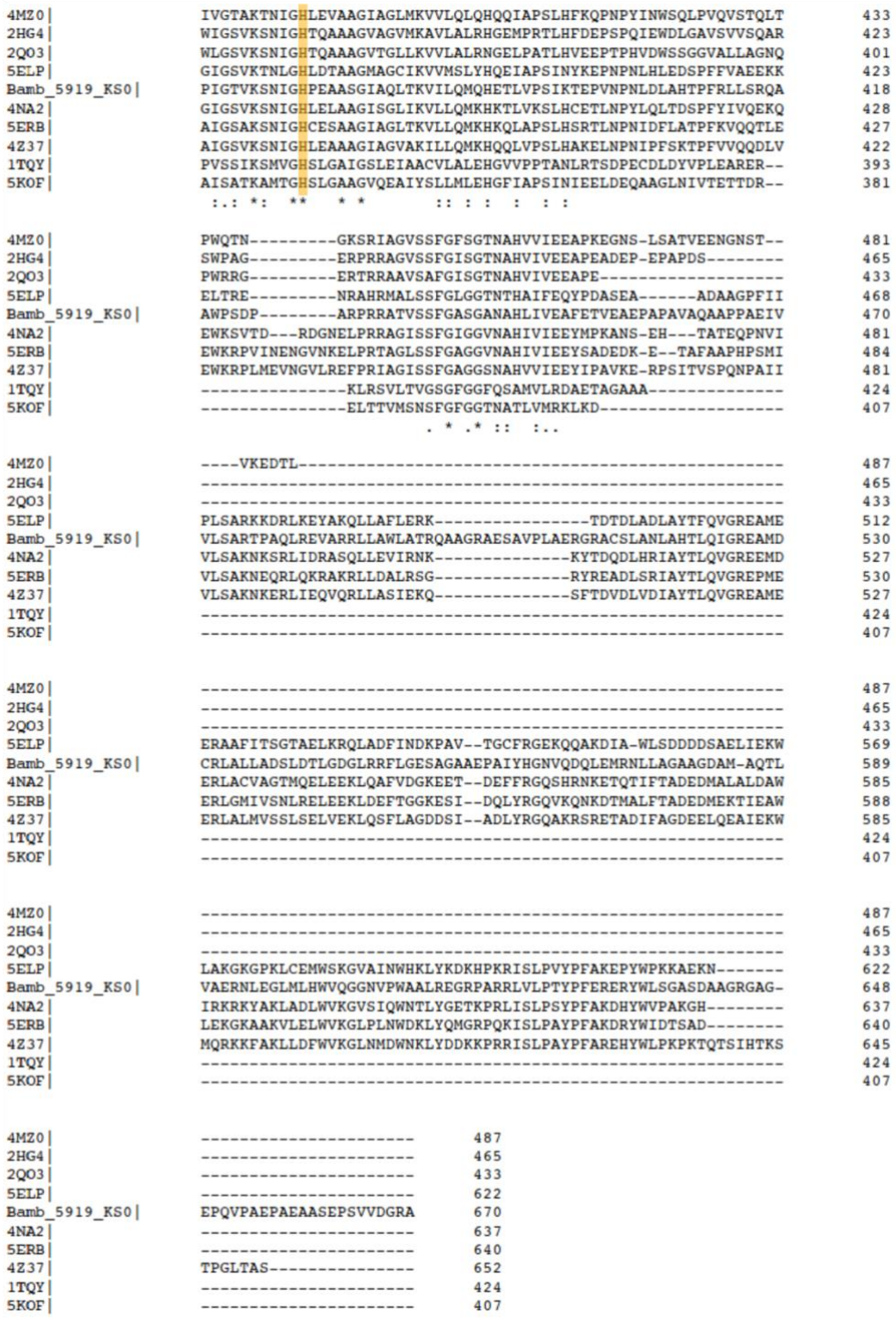


Figure S.4-5 Multiple sequence alignment of Bamb_5919 KS^0, which was used in the project, with the ketosynthase domains with known structures deposited in PDB: 4MZ0 – Moorea producens (Type I PKS), 2HG4 - Saccharopolyspora erythraea (DEBS Module 5, Type I PKS), 2QO3 - Saccharopolyspora erythraea (DEBS Module 3, Type I PKS), 5ELP - Bacillus amyloliquefaciens (Module 1, Type I PKS), 4NA2 – Bacillus subtilis (Type I PKS), 5ERB - Bacillus amyloliquefaciens (Module 5, Type I PKS), 4Z37 - Brevibacillus brevis (Type I PKS), 1TQY - Streptomyces coelicolor (Type II PKS), 5KOF - Escherichia coli (Type II FAS). Bamb_5919_KS0 represents Bamb_5919 KS^0 sequence used in the project. Conserved catalytic triad residues are highlighted in orange.